ABSTRACT

Title of Document: *SALMONELLA* NEWPORT: GENETIC DIVERSITY AND PHYLOGENETIC ANALYSIS

Guojie Cao, Doctor of Philosophy, 2014

Directed By: Professor, Jianghong Meng
Department of Nutrition and Food Science

Salmonella enterica subsp. *enterica* causes over 99% of human salmonellosis. *Salmonella* Newport has ranked in the top three *Salmonella* serotypes associated with foodborne outbreaks in the United States. *S.* Newport is ubiquitous in the environment. *S.* Newport consisted of three lineages. It is necessary to investigate and determine the evolution relationship between *S.* Newport and to identify the genetic diversities of this emerging foodborne pathogen. Whole genome sequencing has played important roles in food safety and public health providing the most accurate information for phylogenetic analysis and more comprehensive picture for comparative genomics.

Total 26 *S.* Newport strains from diverse sources and geographic locations were selected and conducted pyrosequencing to obtain 16-24 × coverage of draft genomes. More than 140,000 SNPs were identified to construct parsimony tree. Phylogenetic analysis indicated that *S.* Newport was divided into two major groups, lineages II and III. Lineage II was further grouped into three subgroups, IIA, IIB, and IIC. Lineage III strains showed

close relationship to each other. Moreover, lineages II and III displayed divergent distance. Comparative genomics identified the region around *mutS* as potential biomarkers to distinguish these two lineages, including *ste* fimbrial operon, transposase, and *cas* genes.

Salmonella pathogenicity islands (SPIs) play essential roles in virulence, metabolism, and host adaptations in *Salmonella*. Due to the significant roles of SPI-5 and SPI-6, the genetic diversities in these two gene clusters may contribute to the various activities in different strains. Both indels and mutations were identified in SPI-5, including two large insertions with over 40 kb encoding phage genes and 146 single nucleotide polymorphisms (SNPs). The phylogenetic tree of SPI-5 genes showed that lineages II and III contained divergent distances. SPI-6 was not identified in Asian strains in subgroup IIA, indicating the potential differences in virulence and host adaptations.

S. Newport multidrug resistant strains have been clinical important issue in the United States. Plasmids contributed to the MDR phenotypes. The common genetic characterizations of these strains could be help to understand the prevalence of MDR strains. In the current study, all MDR strains belonging to one node in IIC and contained unfunctional CRISPR systems.

SALMONELLA NEWPORT: GENETIC DIVERSITY AND PHYLOGENETIC ANALYSIS

By

Guojie Cao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Jianghong Meng, Chair
Dr. Marc Allard
Professor Liangli Yu
Assistant Professor Qin Wang
Professor Charles Mitter

Acknowledgements

First of all, I would like to acknowledge Dr. Jianghong Meng for giving me the opportunity to study in the University of Maryland, College Park. Thanks for Dr. Meng's guidance in science and for his encouragement in life. Dr. Jianghong Meng opens the door of food microbiology and gives a wonderful career path to me. I also want to express my deepest thanks to Dr. Marc Allard and Dr. Eric Brown from Center for Food Safety and Applied Nutrition, Food and Drug Administration guide me to the study of genomics and phylogenetics. Thanks to Dr. Marc Allard, Dr. Liangli Yu, Dr. Qin Wang, and Dr. Charles Mitter for serving on my committee and helping me through numerous discussions and suggestions for my study and dissertation.

I would like to thanks to Dr. Shaohua Zhao from Center for Veterinary Medicine, Food and Drug Administration for her support on my research and encouragement for my future. I also want to thank Dr. Errol Strain, Dr. Robert Stones, Dr. Jie Zheng, Dr. Yan Luo, and Dr. Rebecca Bell for their suggestions and supports on my study and research.

I also want to express my appreciation to Dr. Shuo Wang, my advisor of my Master degree. Dr. Wang supported and encouraged me to study abroad when I studied in Tianjin University of Science and Technology.

I am grateful to all the help and discussions from my previous and current labmates: Dr. Wenting Ju, Magaly Toro, Dr. Lydia Rump, Dr. Xiaodong Xia, Dr. Mohamed Najjar, Dr. Likou Zou, Dr. Qianru Yang, Dr. Yu Hu, Dr. Jinling Shen, Dr. Fei Wang, and Yi Li.

Most important, I want to thank to my parents for their love and understanding during my study. Thank for their support and encouragement. Thanks to my wife Liang Sun for her love, understanding, and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACSSUT	Ampicillin, Chloramphenicol, Streptomycin, Sulfamethoxazole, Tetracycline
BLAST	Basic Local Alignment Search Tool
CDC	Centers for Disease Control and Prevention
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic Acid
FDA	Food and Drug Administration
GIs	Genomic Islands
HGT	Horizontal Gene Transfer
IncA/C	Incompatibility Group A/C
MDR	Multidrug Resistance
MLEE	Multi-locus Enzyme Electrophoresis
MLST	Multi-locus Sequence Typing
MRSA	Methicillin-resistant *Staphylococcus aureus*
MULCLE	Multiple Sequence Comparison by Log-expectation
NGS	Next-generation Sequencing
ORFs	Open Reading Frames
PCR	Polymerase Chain Reaction
PFGE	Pulsed-field Gel Electrophoresis
RNA	Ribonucleic Acid
ncRNA	non-coding RNAs
S. Newport	*Salmonella enterica* subspecies *enterica* serotype Newport

SGI-1	*Salmonella* Genomic Island 1
SNPs	Single Nucleotide Polymorphisms
SPIs	*Salmonella* Pathogenicity Islands
SPI-1	*Salmonella* Pathogenicity Island 1
SPI-5	*Salmonella* Pathogenicity Island 5
SPI5-GI1	*Salmonella* Pathogenicity Island 5 Genomic Island 1
SPI5-GI2	*Salmonella* Pathogenicity Island 5 Genomic Island 2
SPI-6	*Salmonella* Pathogenicity Island 6
SPI6-GI1	*Salmonella* Pathogenicity Island 6 Genomic Island 1
TNT	Tree analysis using New Technology
T3SS	Type III Secretion System
T6SS	Type VI Secretion System
WGS	Whole Genome Sequencing

CHAPTER ONE: LITERATURE REVIEW

Salmonella are gram-negative, facultative anaerobic bacteria belonging to the family *Enterobacteriaceae* (1) and diverged from one common ancestor with *Escherichia coli* approximately 100 million years ago (2). *Salmonella* consists of two different species: *Salmonella bongori* and *Salmonella enterica* (3) including over 2,500 serotypes based on somatic antigens and flagella antigens (1). *Salmonella enterica* subspecies *enterica* composing of over 1,500 serotypes causes most human salmonellosis. *Salmonella enterica* subspecies *enterica* serotype Newport (*S.* Newport) is one emerging and significant serotype associated with foodborne outbreaks in the United States.

Classification of *Salmonella*

The genus *Salmonella* includes two species: *Salmonella bongori* and *Salmonella enterica*. *S. enterica* composes of six subspecies including *enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae*, and *indica* (subspecies I, II, IIIa, IIIb, IV, and VI, respectively). According to the Kauffmann-White serotyping scheme, *Salmonella* also consists of over 2,500 serotypes. The cell surface carbohydrate (somatic antigens or O antigens) and flagella antigens (H antigens) are the two components in the serotyping method (4). Over 1,500 serotypes belong to *S. enterica* subspecies *enterica*, which causes approximate 99% salmonellosis in human and warm blood animals. Moreover, phage typing method differentiates variants in the same serotype, providing a further classification method (5, 6). Moreover, according to host specificity, *Salmonella* includes three categories: 1) those infecting human only such as *S.* Typhi and *S.* Paratyphi A; 2) host adapted ones including *S.* Gallinarum in poultry and *S.* Dublin in cattle; 3) un-adapted ones like *S.* Typhimurium, *S.* Enteritidis and *S.* Newport.

Subtyping methods of *Salmonella*

Different molecular subtyping methods have been used to differentiate close related *Salmonella* including pulse field gel electrophoresis (PFGE) (7, 8), multilocus enzyme electrophoresis (MLEE) (9), multilocus sequence typing (MLST) (7, 10), and clustered regularly interspaced short palindromic repeats (CRISPR) arrays (11, 12). However, these methods do not own enough sensitivity and discriminatory power to differentiate clonal strains (13, 14).

Salmonellosis: an overview

Salmonella causes approximately 1.4 million foodborne illness cases in the United States each year, including 500 deaths (15, 16). Pathogenic *Salmonella* normally enter human body via contaminated water and food. Pets carry *Salmonella* as well, such as reptiles and amphibians (17, 18). Salmonellosis includes typhoidal illness and gastrointestinal illness (19). *S.* Typhi or *S.* Paratyphi A/B causes typhoid fever, a severe systemic human infection (17). Gastrointestinal illnesses includes abdominal pain, vomiting and diarrhea, which are the common symptoms caused by non-typhoidal *Salmonella* (20). The most common four serotypes account for over 50% salmonellosis in the Unites States, including *S.* Typhimurium, *S.* Enteritidis, *S.* Newport, and *S.* Heidelberg (20). *Salmonella* possesses the acid tolerance response to survive in low pH condition in human stomach (21). *Salmonella* enters small intestine and traverse mucous layer to invade nonphagocytic enterocytes of the intestinal epithelium (17, 22, 23).

Outbreaks caused by *S.* Newport

S. Newport causes over 100,000 illness cases annually in the United States (24). *S.* Newport has been named the top three serotypes (9.3%) isolated human sources since

1999 (20). More importantly, infections of *S*. Newport have been on the rise since 1996 (25). The frequency of *S*. Newport in human resource isolates increased 46% in 2009 compared to 1999 in the United States (20). *S*. Newport is an emerging and rising serotype responsible for foodborne outbreaks originated from diverse sources (26-30). For example, one *S*. Newport strain isolated from tomato field caused two multistate outbreaks at 2002 and 2005, respectively. Furthermore, multidrug resistant *S*. Newport has been another emerging public health issue, for both sporadic infections and outbreaks (31-33).

Salmonella pathogenicity islands

Salmonella pathogenicity islands (SPIs) are gene clusters consisting of chromosomal virulence genes and identified only in pathogenic *Salmonella* (19). Virulence genes in SPIs are related to host adaptation, invasion activity, secretion of virulent factors, and survival. SPIs usually are identified next to tRNA with heterogeneous GC content and codon usage, indicating that bacteria acquired SPIs via horizontal gene transfer from foreign genetic sources (19). There are 22 SPIs identified to date. Various serotypes contained distinct distributions of SPIs (34, 35). For example, SPI-7 identified as the largest island to date was determined in *S*. Typhi, but absent in *S*. Typhimurium. Thus, the distributions of SPIs and variations in SPIs would be important indicators to differentiate serotypes or different lineages in the same serotype. SPI-1 to SPI-5 played essential parts for the pathogenicity of *Salmonella*.

SPI-1 is the first identified pathogenicity island with approximate 38.8 kb (36). It was identified in both *Salmonella* species. Thus, *Salmonella* acquired it after the separation with *Escherichia coli*. SPI-1 encodes a type III secretion system, which was involved in

the invasion activity of nonphagocytic cells and proinflammatory responses (36, 37). The ion uptake system (*sitABCD*) in SPI-1 may play important role in the stage of infection when Fe^{2+} and Mn^{2+} become limited. The different GC content of various parts in SPI-1 indicated the acquisition may be one multi step process. Although SPI-1 is the essential one for invasion activity of pathogenic *Salmonella*, it was missing in one clinical strain of *S.* Senftenberg (38). *mutS* was located at the 3' end of SPI-1 and possessed distinct evolution compared to the whole genome (39). Recombination events were not rare in the location around *mutS* (39).

SPI-5 consisted of five genes and was identified in the *S.* Dublin genome first (40). It encodes translocated effector proteins for type III secretion systems (TTSS) in SPI-1 and SPI-2 (41, 42). Mutations in SPI-5 significantly reduced the pathogenicity of *Salmonella* (40).

SPI-6 contained type VI secretion system (T6SS) with diverse functions, which was identified first in *S.* Typhimurium (43-45). The deletion of SPI-6 reduced the invasion activity of *S.* Typhimurium into Hep2 cells (43). Some serotypes carried two SPIs encoding T6SS (SPI-6 and SPI-19), such as *S.* Dublin and *S.* Weltevreden. However, *S.* Agona and *S.* Enteritidis contained only SPI-19. No T6SS was identified in *S.* Virchow and *S.* Paratyphi B (44). Moreover, the *saf* and *tcf* fimbrial operons were identified at downstream of SPI-6.

Evolution of *Salmonella* Newport

S. Newport was considered polyphyletic using MLEE (9) meaning that all strains in one single serotype consisting of several independent lineages in dendrogram (13). MLEE study identified two lineages, divisions I and II, which were related to different hosts.

Most of strains isolated from human belonged to division I and most of those recovered from animals belonged to division II. Multilocus sequence typing (MLST) method also determined the polyphyletic structure of *S.* Newport using three housekeeping genes, *manB*, *fimA*, and *mdh* (46, 47) and identified two groups, both of which related to host sources and multidrug resistance profile (47). Group A strains were mainly isolated from bovine and resistant to different antibiotics, while group B was mainly from avian and sensitive to antibiotics (13, 47). A later MLST study using seven housekeeping genes reported similar results (7). A more recent study (48) including 400 *S.* Newport strains using MLST reported three independent lineages in *S.* Newport population. Most strains from Europe belonged to *S.* Newport I while most of those from North America belonged to *S.* Newport II and III (48). Moreover, approximate 60% MDR strains and all the MDR-AmpC strains belonged to *S.* Newport II. In contrast, 87.5% strains in *S.* Newport III were susceptible ones (48).

The next-generation sequencing technology

Since the first two complete *Salmonella* genomes published at 2001 (34, 35), a total of 36 complete and 383 draft genomes are available in GenBank. The availability of next-generation sequencing (NGS) platforms not only produces huge amount of data but also reduces the cost per genome. The entire genome sequences will provide detailed genetic information of foodborne pathogens and be applied in different research fields, including detection and typing of pathogen, optimize growth and survival strategies, host-pathogen interactions and epidemiology investigations (14, 49). For example, NGS could provide higher sensitivity and discriminatory power to differentiate close related foodborne pathogens belonging to same subtype in traditional molecular subgrouping methods (13,

14), enabling investigators to identify the possible sources of outbreaks in the traceback investigations.

Compared to Sanger sequencing technology, NGS had several advantage characteristics. All NGS platforms do not require cloning step, avoiding biases on the genome data. In contrast, NGS require fragments of sequenced DNA and universal adapters to construct library (50). The NGS library fragments are amplified on a solid surface, one bead or flat glass microfluidic channel, rather than requiring sequencing reaction in microtiter plate wells (50). NGS allows to conduct fragments sequencing and signal detection in parallel with hundreds of thousands to billions of reactions simultaneously. Moreover, certain platform produces longer read length than Sanger technology, facilitating the following assembly process to produce high quality genomes.

Project overview

As *S.* Newport is the important serotype of *Salmonella* causing foodborne outbreaks and is widespread in various foods, it is necessary to investigate the evolution, genetic diversity, multidrug resistance, and virulence factors of *S.* Newport. However, traditional molecular methods could not provide comprehensive picture of this important pathogen. Whole genome sequencing data will not only report the accurate evolutionary history but also shed lights into the genetic diversity. *S.* Newport from diverse hosts and different geographic locations were selected. The objective of current study is to investigate phylogenetic relatedness, genetic diversity of important virulence factors, and possible genetic characteristics of multidrug resistance strains of *S.* Newport. Three specific objectives are as follows:

1) To determine the evolutionary relationship of *S*. Newport from diverse sources and locations. *S*. Newport is widespread and could be isolated from different food products. All previous studies on evolutionary history of *S*. Newport used single gene or small sets of genes, which may not provide accurate information. Thus, it is important to determine the phylogenies among *S*. Newport using whole genome sequencing data.

2) To identify the genetic diversity of *Salmonella* pathogenicity islands in *S*. Newport. SPIs are essential virulence factors playing important roles in invasion activity, survival in host cells, and causing diseases. The presence/absence and mutations in these gene clusters may be responsible for the differences in virulence and could be used as indicator of certain subgroups of *S*. Newport.

3) To determine the genetic characteristics of multidrug resistance strains. Since multi-drug resistance (MDR) is one of the most important medical issues of *S*. Newport, several MDR strains were chosen to explore their potential common genomic background information, which will provide biomarkers for surveillance study and epidemiology investigations.

In the following chapters, three studies are presented for each objective.

References

1. **Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, Petty NK, Kingsley RA, Baumler AJ, Nuccio SP, Contreras I, Santiviago CA, Maskell D, Barrow P, Humphrey T, Nastasi A, Roberts M, Frankel G, Parkhill J, Dougan G, Thomson NR.** 2011. *Salmonella bongori* provides insights into the evolution of the Salmonellae. PLoS Pathog **7:**e1002191.
2. **Doolittle RF, Feng DF, Tsang S, Cho G, Little E.** 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271:**470-477.
3. **Reeves MW, Evins GM, Heiba AA, Plikaytis BD, Farmer JJ, 3rd.** 1989. Clonal nature of *Salmonella* typhi and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. J Clin Microbiol **27:**313-320.
4. **Humphrey T.** 2004. *Salmonella*, stress responses and food safety. Nat Rev Microbiol **2:**504-509.
5. **Hickman-Brenner FW, Stubbs AD, Farmer JJ, 3rd.** 1991. Phage typing of *Salmonella* Enteritidis in the United States. J Clin Microbiol **29:**2817-2823.
6. **De Lappe N, Doran G, O'Connor J, O'Hare C, Cormican M.** 2009. Characterization of bacteriophages used in the *Salmonella enterica* serovar Enteritidis phage-typing scheme. J Med Microbiol **58:**86-93.
7. **Harbottle H, White DG, McDermott PF, Walker RD, Zhao S.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and

antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J Clin Microbiol **44:**2449-2457.

8. **Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW.** 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One **8:**e55254.

9. **Beltran P, Musser JM, Helmuth R, Farmer JJ, 3rd, Frerichs WM, Wachsmuth IK, Ferris K, McWhorter AC, Wells JG, Cravioto A, et al.** 1988. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S.* choleraesuis, *S.* derby, *S.* dublin, *S.* enteritidis, *S.* heidelberg, *S.* infantis, *S.* newport, and *S.* typhimurium. Proc Natl Acad Sci U S A **85:**7753-7757.

10. **Kotetishvili M, Stine OC, Kreger A, Morris JG, Jr., Sulakvelidze A.** 2002. Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. J Clin Microbiol **40:**1626-1635.

11. **Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. Appl Environ Microbiol **77:**1946-1956.

12. **Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C, Weill FX.** 2012. CRISPR

typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One **7:**e36995.

13. **Bell RL, Cao G, Meng J, Allard MW, Keys CE, Hill T, Ottesen AR, Brown EW.** 2012. *Salmonella* Newport Contamination of Produce: Ecological, Genetic, and Epidemiological Aspects, p. In press, *Salmonella*: Classification, Genetics and Disease. Nova Publishers, Hauppauge NY.

14. **Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao G, Meng J, Stones R.** 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med **364:**981-982.

15. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-625.

16. **Voetsch AC, Van Gilder TJ, Angulo FJ, Farley MM, Shallow S, Marcus R, Cieslak PR, Deneen VC, Tauxe RV.** 2004. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin Infect Dis **38 Suppl 3:**S127-134.

17. **Haraga A, Ohlson MB, Miller SI.** 2008. Salmonellae interplay with host cells. Nat Rev Microbiol **6:**53-66.

18. **Woodward DL, Khakhria R, Johnson WM.** 1997. Human salmonellosis associated with exotic pets. J Clin Microbiol **35:**2786-2790.

19. **Sabbagh SC, Forest CG, Lepage C, Leclerc JM, Daigle F.** 2010. So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. FEMS Microbiol Lett **305:**1-13.

20. **CDC.** 2009. *Salmonella* Surveillance: Annual Summary. Atlanta, Georgia: US Department of Health and Human Services, CDC.

21. **Garcia-del Portillo F, Foster JW, Finlay BB.** 1993. Role of acid tolerance response genes in *Salmonella* Typhimurium virulence. Infect Immun **61:**4489-4492.

22. **Michetti P, Mahan MJ, Slauch JM, Mekalanos JJ, Neutra MR.** 1992. Monoclonal secretory immunoglobulin A protects mice against oral challenge with the invasive pathogen *Salmonella* Typhimurium. Infect Immun **60:**1786-1792.

23. **Francis CL, Starnbach MN, Falkow S.** 1992. Morphological and cytoskeletal changes in epithelial cells occur immediately upon interaction with *Salmonella* Typhimurium grown under low-oxygen conditions. Mol Microbiol **6:**3077-3087.

24. **Greene SK, Daly ER, Talbot EA, Demma LJ, Holzbauer S, Patel NJ, Hill TA, Walderhaug MO, Hoekstra RM, Lynch MF, Painter JA.** 2008. Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. Epidemiol Infect **136:**157-165.

25. 2011. Vital signs: incidence and trends of infection with pathogens transmitted commonly through food--foodborne diseases active surveillance network, 10 U.S. sites, 1996-2010. MMWR Morb Mortal Wkly Rep **60:**749-755.

26. **Van Beneden CA, Keene WE, Strang RA, Werker DH, King AS, Mahon B, Hedberg K, Bell A, Kelly MT, Balan VK, Mac Kenzie WR, Fleming D.** 1999. Multinational outbreak of *Salmonella enterica* serotype Newport infections due to contaminated alfalfa sprouts. JAMA **281:**158-162.

27. **Sivapalasingam S, Barrett E, Kimura A, Van Duyne S, De Witt W, Ying M, Frisch A, Phan Q, Gould E, Shillam P, Reddy V, Cooper T, Hoekstra M, Higgins C, Sanders JP, Tauxe RV, Slutsker L.** 2003. A multistate outbreak of *Salmonella enterica* Serotype Newport infection linked to mango consumption: impact of water-dip disinfestation technology. Clin Infect Dis **37:**1585-1590.

28. **Kirk MD, Little CL, Lem M, Fyfe M, Genobile D, Tan A, Threlfall J, Paccagnella A, Lightfoot D, Lyi H, McIntyre L, Ward L, Brown DJ, Surnam S, Fisher IS.** 2004. An outbreak due to peanuts in their shell caused by *Salmonella enterica* serotypes Stanley and Newport--sharing molecular information to solve international outbreaks. Epidemiol Infect **132:**571-577.

29. **Schneider JL, White PL, Weiss J, Norton D, Lidgard J, Gould LH, Yee B, Vugia DJ, Mohle-Boetani J.** 2011. Multistate outbreak of multidrug-resistant *Salmonella* newport infections associated with ground beef, October to December 2007. J Food Prot **74:**1315-1319.

30. **Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ.** 2013. Outbreak-associated *Salmonella enterica* Serotypes and Food Commodities, United States, 1998-2008. Emerg Infect Dis **19:**1239-1244.

31. **Zansky S, Wallace B, Schoonmaker-Bopp D, Smith P, Ramsey F, Painter J, Gupta A, Kalluri P, Noviello S.** 2002. From the Centers for Disease Control and

Prevention. Outbreak of multi-drug resistant *Salmonella* Newport--United States, January-April 2002. JAMA **288:**951-953.

32. **Varma JK, Marcus R, Stenzel SA, Hanna SS, Gettner S, Anderson BJ, Hayes T, Shiferaw B, Crume TL, Joyce K, Fullerton KE, Voetsch AC, Angulo FJ.** 2006. Highly resistant *Salmonella* Newport-MDRAmpC transmitted through the domestic US food supply: a FoodNet case-control study of sporadic *Salmonella* Newport infections, 2002-2003. J Infect Dis **194:**222-230.

33. **Gupta A, Fontana J, Crowe C, Bolstorff B, Stout A, Van Duyne S, Hoekstra MP, Whichard JM, Barrett TJ, Angulo FJ.** 2003. Emergence of multidrug-resistant *Salmonella enterica* serotype Newport infections resistant to expanded-spectrum cephalosporins in the United States. J Infect Dis **188:**1707-1716.

34. **Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848-852.

35. **McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller**

W, Stoneking T, Nhan M, Waterston R, Wilson RK. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature **413:**852-856.

36. **Mills DM, Bajaj V, Lee CA.** 1995. A 40 kb chromosomal fragment encoding *Salmonella* typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome. Mol Microbiol **15:**749-759.

37. **Galan JE.** 1999. Interaction of *Salmonella* with host cells through the centisome 63 type III secretion system. Curr Opin Microbiol **2:**46-50.

38. **Hu Q, Coburn B, Deng W, Li Y, Shi X, Lan Q, Wang B, Coombes BK, Finlay BB.** 2008. *Salmonella enterica* serovar Senftenberg human clinical isolates lacking SPI-1. J Clin Microbiol **46:**1330-1336.

39. **Brown EW, Mammel MK, LeClerc JE, Cebula TA.** 2003. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens. Proc Natl Acad Sci U S A **100:**15676-15681.

40. **Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE.** 1998. Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. Mol Microbiol **29:**883-891.

41. **Hensel M.** 2004. Evolution of pathogenicity islands of *Salmonella enterica*. Int J Med Microbiol **294:**95-102.

42. **Knodler LA, Celli J, Hardt WD, Vallance BA, Yip C, Finlay BB.** 2002. *Salmonella* effectors within a single pathogenicity island are differentially expressed and translocated by separate type III secretion systems. Mol Microbiol **43:**1089-1103.

43. **Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Lofdahl S.** 1999. Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. Mol Microbiol **33:**612-622.

44. **Blondel CJ, Jimenez JC, Contreras I, Santiviago CA.** 2009. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. BMC Genomics **10:**354.

45. **Jani AJ, Cotter PA.** 2010. Type VI secretion: not just for pathogenesis anymore. Cell Host Microbe **8:**2-6.

46. **Sukhnanand S, Alcaine S, Warnick LD, Su WL, Hof J, Craver MP, McDonough P, Boor KJ, Wiedmann M.** 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella* enterica serotypes. J Clin Microbiol **43:**3688-3698.

47. **Alcaine SD, Soyer Y, Warnick LD, Su WL, Sukhnanand S, Richards J, Fortes ED, McDonough P, Root TP, Dumas NB, Grohn Y, Wiedmann M.** 2006. Multilocus sequence typing supports the hypothesis that cow- and human-associated *Salmonella* isolates represent distinct and overlapping populations. Appl Environ Microbiol **72:**7575-7585.

48. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010. Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

49. **Begley M, Hill C.** 2010. Food safety: What can we learn from genomics? Annu Rev Food Sci Technol **1:**341-361.

50. **Mardis ER.** 2013. Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto Calif) **6:**287-303.

CHAPTER II: PHYLOGENETICS AND DIFFERENTIATION OF *SALMONELLA* NEWPORT LINEAGES USING WHOLE GENOME SEQUENCING

Abstract

Salmonella enterica subspecies *enterica* serotype Newport (*S*. Newport) has ranked in the top three *Salmonella* serotypes associated with foodborne outbreaks from 1995 to 2011 in the United States. In the current chapter, a total of 26 *S*. Newport strains isolated from diverse hosts and geographic locations were selected and conducted 454 shotgun pyrosequencing procedures to obtain 16-24 × coverage of high quality draft genomes. Comparative genomics analysis of 28 *S*. Newport strains (including 2 reference genomes) and 15 outgroup genomes identified over 140,000 informative SNPs. A resulting phylogenetic tree consisted of four subgroups indicating that *S*. Newport had a clear geographic structure. Strains from Asia were divergent from those from the Americas. Our findings demonstrated that analysis using whole genome sequencing data resulted in a more accurate picture of phylogeny compared to those using single genes or small sets of genes. I selected location around *mutS* in *S*. Newport to differentiate lineages, including region between *invH* and *mutS* at the 3' end of *Salmonella* pathogenicity island 1 (SPI-1), *ste* fimbrial operon, and clustered regularly interspaced short palindromic repeats (CRISPR) associated-proteins (*cas*). These genes in the outgroup genomes held high similarities with either *S*. Newport lineage II or III at the same locus. *S*. Newport lineages II and III owned different evolutionary histories in this region. Our data

demonstrated genetic flow and homologous recombination events around *mutS*. *S.* Newport lineages II and III diverged early in the serotype evolution and have evolved largely independently. Moreover, genes that could delineate subgroups within the phylogenetic tree and that could be used as potential biomarkers for trace-back investigations during outbreaks were determined. Thus, whole genome sequencing data provides better understanding of genetic background of pathogenicity and evolutionary history of *S.* Newport and also identified additional markers for epidemiological response.

Introduction

Salmonellosis is a major contributor to global public health burden. In the United States, non-typhoid *Salmonella* cause an estimated 1.4 million gastroenteritis cases (1) and several billion dollars of economic loss annually (2). Non-typhoid *Salmonella* accounts for only 11% of foodborne illnesses (3), whereas causes 35% of hospitalizations and 28% of the deaths related to foodborne illnesses (4). Over 1,500 serotypes belong to *Salmonella. enterica* subsp. *enterica* (5). *Salmonella enterica* subspecies *enterica* serotype Newport (*S.* Newport) ranked in the top three *Salmonella* serotypes associated with foodborne outbreaks in the United States. (3). The number of *S.* Newport outbreaks increased markedly since 1995, causing at least 100,000 infections annually (3). *S.* Newport was responsible for several multistate outbreaks associated with tomatoes, ground beef, alfalfa sprouts, and other food products since 2002 (3, 6-9). *S.* Newport displayed high levels of genomic diversity and possessed polyphyletic structure according to multilocus enzyme electrophoresis (MLEE) (10) and multilocus sequence typing (MLST) (11-13). *S.* Newport contained three lineages according to MLST analysis (14). Most strains from Europe belong to *S.* Newport lineage I, whereas most strains from North America belong to lineages II and III (14).

Recombination events played a key role in the evolution of *Salmonella* (15, 16). Brown et al. (17) indicated that evolution of *mutS* was distinct from the whole genome, and recombination events were not rare at the loci around *mutS* including 3' end region of SPI-1, *ste* fimbrial operon and *cas* gene cluster. SPI-1 is a 40 kb gene cluster encoding type III secretion system (T3SS) (18). It was identified in both *Salmonella enterica* and

Salmonella bongori, although one study reported that an *S.* Senftenberg clinical strain did not possess SPI-1 (19).

Clustered regularly interspaced short palindromic repeats (CRISPR)/*cas* systems are present in 90% archaea and approximately 40% of bacteria (20, 21) and considered important immune system to protect bacteria against foreign genetic elements as well as to help microbes to survive phage predation; the CRISPR/cas system also facilitated the microbes to adapt to specific niche (22-24). The diversity of CRISPR/*cas* system in *E. coli* and *Salmonella* revealed phylogeny of the *cas* protein family in different serotypes (25).

Since the first two *Salmonella* whole genome sequences were available in 2001 (26, 27), a total of 36 complete and 383 draft genomes were released in GenBank, including *S.* Newport SL254 and SL317 besides our data. Whole genome sequencing has been increasingly used as a tool for evolutionary studies and epidemiological investigations (9, 28-32). In the current chapter, we performed pyrosequencing to obtain 16-24 × coverage high quality draft genomes of 26 *S.* Newport strains from diverse sources and geographic locations. Our data demonstrated the phylogenetic relationship among *S.* Newport strains and revealed variations around *mutS* gene, providing genetic evidence of recombination events. Moreover, genes delineating major lineages and subgroups were identified and could be used as biomarkers to develop tools for trace-back studies for epidemiology and outbreak investigations. This chapter was published (33).

Materials and Methods

Bacterial strains

A total of 26 *S.* Newport strains isolated from diverse hosts and geographic locations were selected (TABLE II-1). *S.* Newport SL254 (ABEN01000000) and *S.* Newport SL317 (ABEW00000000) were downloaded from GenBank as reference genomes. Other 15 *Salmonella* genomes of different serotypes were chosen to be outgroups according to pervious study (34, 35). They are *S.* I 4,[5],12:i- SL474 (ABAO00000000), *S.* Kentucky CDC191 (ABEI00000000), *S.* Kentucky CVM29188 (ABAK00000000), *S.* Dublin CT_02021853 (CP001144), *S.* Gallinarum 287/91 (AM933171), *S.* Tennessee CDC07-0191 (ACBF00000000), *S.* Typhimurium 14028S (NC_016856.1), *S.* Typhimurium LT2 (NC_003197.1), *S.* Typhimurium SL1344 (NC_016810.1), *S.* Typhimurium D23580 (NC_016854), *S.* Choleraesuis SC-B67 (AE017220), *S.* Paratyphi C RKS4594 (CP000857), *S.* Virchow SL491 (ABFH00000000), *S.* Saintpaul SARA29 (ABAN00000000) and *S.* Hadar RI_05P066 (ABFG00000000).

Pulsed field gel electrophoresis (PFGE)

PFGE was performed according to the procedure as previously described (11).

Genome sequencing, assembling and annotation

Bacterial cells were pelleted from one ml of pure tryptic soytone broth from overnight culture by centrifugation and DNA were prepared using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. I sequenced 26 *S.* Newport strains using Roche 454 GS-FLX Titanium sequencer (Roche, Branford, CT) to obtain 16-24 × coverage of draft genomes (except strain canine_AZ_2003 with 9 ×

coverage). This platform provides longer read lengths than other nest generation sequencing platforms. *De novo* assemblies were performed using the Roche Newbler (v 2.3) software package. Annotation of resulting contigs was finished by NCBI according to Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (36). Phylogenetically informative SNPs were identified using two independent alignment methods: 1) multiple genome alignment of whole genome sequencing contigs using Mauve (37), and 2) clustering of annotated open reading frames (ORFs) using reciprocal best Basic Local Alignment Search Tool (BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi) hits with a 70% sequence identity setting followed by alignment with multiple sequence comparison by log-expectation (MULCLE) (38).

Phylogenetic tree construction

Parsimony phylogenetic tree was constructed based on 147,780 concatenated informative SNPs using tree analysis using new technology (TNT) (39) with finding minimum tree length 20 times and 100,000 iterations. Seven housekeeping genes were extracted to perform MLST analysis. Concatenated housekeeping gene sequences were analyzed using TNT (39) with same parameters. Moreover, we performed multiple sequence alignment using MULCLE (38) in SeaView 4 (40) and identified concatenated sequences of *cas* genes (*cas1*, *cas2*, *cas5*, *cse1*, *cse2*, *cse3* and *cse4*) with around 6 kb. Strains frog_Vietnam, fish_Hong_Kong, fish_Vietnam, canine_AZ_2003 and pig_ear_CA were not involved in this analysis. I performed TNT to display evolutionary relatedness of *cas* genes with the same parameters.

Recombination analysis

ClonalFrame was used (41) to analyze effects of recombination events on the evolutionary history of *S.* Newport. *S.* DublinCT_02021853 was used as an outgroup genome to display the recombination events and substitutions between *S.* Newport lineages II and III, which showed close relatedness to both lineages. All 29 *Salmonella* genomes were aligned using progressiveMauve (37) with the default settings. I used the stripSubsetLCBs (locally collinear blocks) (http://gel.ahabs.wisc.edu/mauve/snapshots/) script to extract core blocks, which created core alignments longer than 500 bp that included all 29 genomes. I obtained total 510 LCBs. Given the computational demands necessary to analyze all 510 blocks simultaneously, we created three separate datasets each consisting of 50 randomly selected blocks. I ran ClonalFrame (41) on each of these three datasets with estimated parameters based on 200,000 generations of which the first 100,000 generations served as burn-in. The thinning interval was set to 100. The Gelmin-Rubin statistic was used to determine whether the independent runs had converged on similar parameter estimates, which also provided evidence that random subsets of the genome did not bias our results. Furthermore, Mauve (37) was used to compare the genomic organizations.

Differences of gene cluster between *invH* and *mutS*

I performed blastp to search best match of genes between *invH* and *mutS*, including gene clusters 1, 2, and 3. Tblastn was employed to verify the searching results.

Pairwise distance matrix

MEGA 5.05 (42) was employed to calculate evolutionary distance (no. of differences) over sequence pairs between groups with 1,000 bootstrap iterations.

Searching for most variable genes

Custom software was employed to look for the genes and informative SNPs defining the major lineages and subgroups. This was a GUI shell around open source software. In this analysis, a total of 29 genomes including all 28 *S.* Newport strains and *S.* Choleraesuis SC-B67 as an outgroup genome were selected. UCLUST algorithm (43) was employed to search gene families, using default settings with a 95% sequence identities cutoff. Maximum and minimum length of a gene cluster to search was 58,000 and 10 bp, respectively. MUSCLE (38) was employed to perform alignment with default settings. SNPs of these gene clusters were detected and were used to create phylogenetic matrix to construct phylogenetic tree using TNT (39) and count the informative SNPs that delineating major and subgroups on the nodes. Then we selected the genes containing the largest number of informative SNPs that defined major lineages and subgroups.

Results

Phylogenetic Relationship

Pyrosequencing was used to obtain 16-24 × coverage (except strain from canine_AZ_2003 with 9 × coverage) of high quality draft genomes of 26 *S.* Newport strains with genome sizes ranging from 4.6M bp to 5.0M bp (TABLE II-1). Other 15 genomes were selected as outgroup genomes to determine evolutionary relatedness and test polyphyly with *S.* Newport according to previous studies (34, 35) and one unpublished study of Center for Food Safety and Applied Nutrition, FDA. The outgroup genomes had close relatedness with *S.* Newport or were able to separate *S.* Newport. *S.* Newport SL254 and *S.* Newport SL317 were selected as reference genomes of *S.* Newport lineages II and III, respectively (14). *S.* Newport strains farm_1_VA_2007 and farm_15_VA_2007 were environmental isolates from a farm on the Virginia Eastern Shore. Among the 26 draft genomes, the largest genome size was 5.01M bp of cannine_AZ_2003, while the smallest one was 4.65M bp of pepper_Vietnam. There was no correlation between genome size and major lineages or subgroups.

Over 147,000 informative SNPs were identified from multiple genome alignment and were used to construct a parsimony phylogenetic tree (FIG II-1). All 28 *S.* Newport genomes (including *S.* Newport SL254 and SL317) were grouped into two major lineages (FIG II-1), lineages II and III (14). *S.* Newport lineage II was further divided into subgroups IIA, IIB and IIC. *S.* Newport displayed a clear geographic structure. For example, strains frog_Vietnam, fish_Hong_Kong, fish_Vietnam, shrimp_India, squid_Vietnam and pepper_Vietnam were placed in two subgroups (IIA and IIB) in lineage II, and divergent from those from the Americas (IIC). The two Vietnamese strains

in subgroup IIA grouped together to the exclusion of the other Asian strain and the same grouping of Vietnamese strains was also seen in IIB. IIC included one Mexican strain (cheese_Mexico) and many North American strains and defined the Americas clade. However, this structure was imperfect with pig_ear_CA located in IIA, an otherwise Asian clade. The U.S. strains from various hosts were diverse and grouped into both major lineages. All strains in lineage III were isolated from the United States. *S.* Newport lineages II and III were polyphyletic, namely, lineage III displayed closer evolutionary relationship with *S.* Hadar and *S.* Typhimurium outgroups than lineage II (FIG II-1).

Since MLST has been used as a common analysis tool to study the phylogenetic relatedness and epidemiology of *Salmonella*, we extracted seven housekeeping genes (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*) in each genome (except strain canine_AZ_2003 because of sequence quality) and performed MLST analyses (FIG II-2). MLST indicated that lineage II was grouped into three subgroups with minor differences with the genome phylogenetic tree. For example, subgroup IIA showed closer relatedness with IIB than IIC. Additionally, lineages II and III were separated by outgroup genomes, although outgroups displayed different relatedness compared with the genome tree (FIG II-1). For example, *S.* Virchow, *S.* Paratyphi C and *S.* Choleraesuis showed closer relationship with lineage II.

Furthermore, we listed pairwise SNP variation among these four subgroups (IIA, IIB, IIC and III) and five different outgroups (TABLE II-2). The inter-lineage SNP diversity was remarkable indicating the extensive genomic diversity in *S.* Newport population. For example, the distance between lineage III and subgroup IIB was approximately 36,800 SNPs, which was greater than the one between lineage III and *S.* Hadar RI_05P066

(approximately 34,400). In lineage II, subgroup IIC had closer relationship with IIB than IIA (FIG II-1, TABLE II-2).

I analyzed informative SNPs that define subgroups in the phylogenetic trees (TABLE II-3). The SNPs that delineated the subgroups originated from various regions around the genome of *S*. Newport and included a variety of genes assigned to diverse functions including virulence, DNA replication and repair, and metabolism. For example, there were approximately 13,000 informative SNPs that changed only once and could differentiated lineages II and III. Additionally, there are 2831, 2508, 1259 informative SNPs that defining subgroups IIA, IIB, IIC, respectively.

Moreover, we listed variable genes delineating various subgroups with their SNP changes, gene names and genome locus alignment coordinates (TABLE II-3). I selected informative SNPs from 20 most variable genes (with highest number of informative SNPs that changed once and defined all members of each major lineage and subgroup) defining the two major lineages subgroups in lineage II. For example, *tpiA* gene (SNSL254_A4410) could be used as a marker to differentiate lineages II and III. At position 91 of the alignment, nucleotides in *S*. Newport SL254 and SL317 were A and G, respectively, and amino acid changed from threonine to alanine. Variable genes found within subgroups IIA and IIB could be used as markers for Asian strains. Furthermore, the most variable genes with the largest numbers of informative SNPs could be used as targets of resequencing (TABLE II-3). For example, there were 78 SNPs in *carB* (3228 bp) and 71 SNPs in *aceE* (2664 bp).

A cluster of multidrug-resistant *S*. Newport strains was placed in IIC, namely, node M in the parsimony tree (FIG II-1). Previous studies indicated that *S*. Newport MDR-AmpC

(resistant to third generation cephalosporins containing an ampC ß–lactamase gene (44, 45)) strains belonged to lineage II (14). I identified the 20 most variable genes containing informative SNPs defining this subgroup (TABLE II-3). Interestingly, *acrD* (SNSL254_A2674), encoding a multidrug efflux protein, contained 12 informative SNPs defining subgroup IIC with 3114 bp length. All strains of subgroup IIC had nucleotide C at the position 84, whereas all other *S*. Newport strains had T at the same position.

Additionally, five MDR strains in subgroup IIC were grouped together, including cattle_AZ_2003, *S*. Newport SL254, cattle_NC_2003, cannine_AZ_2003, ground_beef_GA_2004 (FIG II-1, TABLE II-1). The one exception was swine_IL_2001, which was MDR but separated from the other MDR strains by ground_turkey_NM_2008 that was only resistant to Tetracycline. Furthermore, we analyzed the informative SNPs that delineating the node M (FIG II-1) and determined 33 informative SNPs (TABLE II-3). It is notable that 17 of the informative SNPs were non-synonymous. For example, in *ksgA* encoding RNA dimethyltransferase associated with antibiotic resistance (46), the node M carried A at the position of 689, whereas the other *S*. Newport strains possessed G at the same position (TABLE II-3), resulting in an amino acid change from asparagine to glycine.

I compared the genomic organization of *S*. Newport in distinct subgroups (FIG II-3). I used *S*. Newport SL254 as reference genome, which was complete genome in IIC, to compare with *S*. Newport SL317, pig_ear_CA and fish_Vietnam belonging to lineage III, subgroups IIA and IIB, respectively. Our data indicated that large indels and rearrangement events could be found, although the general genomic organizations are same.

Because of the importance of recombination events in the evolution of *S*. Newport, we performed ClonalFrame (41) analyses to reveal the effects of recombination events on the evolutionary history of *S*. Newport (FIG II-4). Our data indicated that the r/m (FIG II-4A: r/m equals the ratio of possibilities that a given site is altered through recombination event and substitution) and ρ/θ (FIG II-4B: ρ/θ equals the ratio of rates of recombination event and substitution occur at a locus) ratios were 1.68 and 0.1, respectively. Moreover, the genomic representation mode was selected to display the recombination events (red line) and substitution (green triangle) on the nodes of II vs. III and IIA vs. IIB&C (FIG II-4 C and D). Our data indicated that recombination events in lineage II happened more frequently than those between lineages II and III.

Region between *invH* and *mutS*

Genes between *invH* and *mutS* in both lineages displayed distinct contents and were defined as gene clusters 1 and 2, respectively (TABLE II-4). Because they were conserved in each lineage, *S*. Newport strains SL254 and chicken_MO were selected for further analysis of these gene clusters. All 15 outgroup genomes possessed gene cluster 1 at the same location with minor differences were identified.

Gene cluster 1 contained six genes ranging from 282 to 669 bp and encoding ABC transport system protein, transposase, phosphatase and membrane protein (TABLE II-4). The GC content ranged from 41.7 to 55.9%. The best hits of blastp against these genes belonged to various serotypes. For example, the best blast match of *pphB* gene in gene cluster 1 was *S.* Typhimurium LT2; and the best blastp match of *tnp* gene encoding a transposase in *S*. Newport SL254 was *Enterobacter cloacae* with 84% identities and 95% coverage. *S*. Typhimurium and *S*. Bardo had 62% identities and 95% coverage of *tnp*

with *S*. Newport SL254. Blast matches of other genes in gene cluster 1 were distributed broadly across *Salmonella* serotypes. Additionally, one large insertion (gene cluster 3) was identified at the 3' end of gene cluster 1 in strain fish_Hong_Kong (FIG II-1, TABLE II-5), including genes encoding transposase, integrase, phage related proteins and proteins of type I restriction modification system in *Vibrio*. For example, according to blastp search *hsdS* gene showed 61% identities and 79% positives to gene in *Vibrio splendidus*; *hsdM* gene showed 84% identities and 92% positives to gene in *Vibrio metschnikovii*. These findings suggested that region between *invH* and *mutS* was hot spot for horizontal gene transfer or recombination events and could facilitate acquisitions of new genetic elements.

A total of six genes ranging from 90 to 738 bp (46.7 to 55% GC contents) were identified in gene cluster 2 in strain chicken_MO (TABLE II-4). The best blastp hits of the genes, except *tnpA* gene, were *S*. Newport SL317; and the best blastp match of *tnpA* was *S*. Dublin CT_02021853. The best blastp hit of *insF* in chicken_MO was a transposase of *S*. Newport SL317 with 100% identities and 100% coverage, and *insF* was also found in *S*. arizonae 62:z4,z23 with 88% identities and 68% coverage and in *S*. Hadar RI_05P066 with 93% identities and 100% coverage, but not in other *Salmonella* serotypes. The other four genes in gene cluster 2 had a broader distribution among different serotypes of *Salmonella*. The best blastp hits of these four genes were proteins from *S*. Newport SL317 and *S*. Dublin CT_02021853 with at least 93% identities and with 100% coverage. Additionally, *tnpA* gene in chicken_MO was absent in *S*. Newport lineage II and *S*. Virchow SL491 but present in all other genomes in the current study.

Gene cluster encoding fimbrial operon and *cas* genes

Similar to those between *invH* and *mutS*, genes at the 3' end of *mutS* displayed significant variations between lineages II and III, and genes of lineage II showed high similarity with those of the outgroup genomes (TABLE II-6). *steABCDEF* fimbrial operon located between *relA* and *mazG* was identified in lineage II and all outgroup genomes, but not lineage III. Blastp results demonstrated that this fimbrial operon was present in certain *Salmonella* serotypes. Lineage III strains had only two genes at the same locus, encoding RelE/ParE family plasmid stabilization system protein (SNSL317_A4074) and putative addiction module antidote protein (SNSL317_A4073). Interestingly, *steF* in *S.* Newport SL254 and genes between *relA* and *mazG* in *S.* Newport SL317 were found adjoining each other in *S.* Typhi CT18.

I defined *cas* genes located at the 3' end of *mutS* in lineages II and III as *cas* sequence 1 and 2, respectively. Sequence alignment showed significant variations between these two sequences. Moreover, there are four collapsing groups in the parsimony tree because the sequence identities were almost 100% in each group, respectively. For example, there was only one substitution found at position 333 of strain from ground_turkey_MD_2003 in total 5,781 bp compared with other four sequences in the group (data not shown). A parsimony tree was generated based on *cas* alignments (FIG II-5), showing that *cas* proteins in lineages II and III displayed divergent phylogenetic relatedness, and were separated by outgroup genomes. For example, *cas* genes of *S.* Paratyphi C and *S.* Choleraesuis displayed closer relatedness with *cas* sequence 2 of lineage III than lineage II strains (FIG II-5).

Discussion

In the current chapter, whole genome sequencing data revealed that *S.* Newport lineages II and III were polyphyletic to each other and were separated by other *Salmonella* serotypes, such as *S.* Hadar and *S.* Typhimurium (FIG II-1). A phylogenetic tree from previous study (16) based on whole genome sequencing data suggested similar relationships and illustrated that *S.* Virchow SL491 phylogenetically displayed closer relationship with *S.* Newport lineages than others. In Fricke's (16) study, 28 sequenced genomes representing 21 serotypes of *S. enterica* were selected to demonstrate the evolutionary history of subgroups of *S. enterica*. Conversely, we focused on variability between major lineages and subgroups of *S.* Newport with 15 outgroup genomes. Although a phylogenetic tree based on whole genome sequencing data provided a more accurate dendrogram than traditional subtyping methods (32), sampling was a critical factor to accomplish study research goals. Importantly, these findings provided an insightful picture to reconstruct the evolutionary history of *S.* Newport. As more sequenced *Salmonella* genomes become available, more accurate and comprehensive phylogenetic information will give a better understanding of the evolution and ecology of *Salmonella*, for both subspecies and single serotypes (47).

Conventional subtyping methods, such as MLST and PFGE, have been used to differentiate pathogenic strains during outbreaks and trace-back investigations and to study the phylogenetic organization of pathogens. MLST analyses (FIG II-2) indicated that *S.* Newport was divided into two major lineages and was separated by outgroup genomes. Lineage II was divided further into three subgroups; however, subgroup IIA displayed closer relatedness with IIC than IIB, which was different with the whole

genome based parsimony tree (FIG II-1). This was not unexpected as the genomic database was significantly larger than the MLST database. MLST indicated that these seven housekeeping genes were valuable to differentiate major and subgroups of *S.* Newport, though MLST may not accurately show the relationships in different subgroups. Therefore, whole genome sequencing data was able to provide more accurate phylogenetic relationship than small sets of genes. However, PFGE often may not be able to differentiate highly clonal strains (29, 32). A combination of whole genome sequencing and phylogenetic analysis has been proven to provide enough accuracy and sensitivity for epidemiological investigations (29, 32). In the current study, PFGE was not able to delineate the major lineages correctly, as expected (FIG II-1 and II-6). For example, according to the PFGE profile (FIG II-6), subgroup IIB strains were located with strains of lineage III to form a lineage unsupported by sequence analysis. Bell et al. (9) demonstrated that whole genome sequencing and phylogenetic analysis were able to differentiate *S.* Newport strains with an identical PFGE pattern during an outbreak case study, providing detailed information about *S.* Newport's complex ecology to the investigators.

The parsimony tree had a clear geographic structure, which appears to be a common characteristic of *Salmonella* (29, 32). *S.* Newport strains isolated from Asia were grouped together and divergent from those from the Americas (FIG II-1). Lineages II and III displayed extensive genomic diversity. For example, lineage II strains from North America had closer evolutionary relatedness with those of lineage II from Asia than ones of lineage III from North America, suggesting that the geographic structure could be observed only among highly clonal lineages, but may not be apparent among the major

lineages. Moreover, there was a diverse phylogenetic structure with strain cheese_Mexico genetically unique to other strains of North America (subgroup IIC in FIG II-1), suggesting that *S.* Newport strains isolated from different states in the United States and in the Americas may have finer geographic structure. *S.* Newport strains from Asia showed diverse geographic structure. Strains from Vietnam in subgroups IIA and IIB displayed closer relationship than with another Asian strain within the same subgroup (FIG II-1). In addition, strains from Vietnam originated from different subgroups of lineage II, such as frog_Vietnam and squid_Vietnam. However, pig_ear_CA was located in subgroup IIA, which was otherwise composed of Asian strains (FIG II-1). This result indicated that *S.* Newport strains from Asia or this subgroup may have extensive genomic diversity and that geographic structure may be better identified among the most highly clonal lineages. I hypothesized that strain pig_ear_CA may be related to a food import or export from the Pacific Rim. Analysis of more isolates is needed to confirm the pig-ear subgroup.

The SNPs that delineating each subgroup (TABLE II-3) were the most valuable for both targeted resequencing efforts and rapid subtyping methods for trace-back of future *S.* Newport outbreak investigations and diagnosis, including the SNPs defining MDR strains, though plasmids likely play a critical role for antibiotic resistance of *S.* Newport (48, 49). Sangal et al. (14) indicated that most of lineage III strains were pan-susceptible and all MDR-AmpC strains were exclusively associated with two sequence types (STs) in lineage II. Similarly, all MDR strains in the present study were grouped together in subgroup IIC. I hypothesized that the plasmids of MDR strains in the present study had the same backbone as *Y. pestis* plP1202 and *S.* Newport pSN254, which could be broadly

disseminated among MDR pathogens via horizontal or vertical gene transfer (48). Genes associated with antibiotic resistance, *acrD* and *ksgA*, delineated subgroup IIC and node M (FIG II-1), respectively. Matsumura et al. (50) suggested that *acrD* contributed significantly to the formation of biofilm of *E coli* K-12. AcrD also played a major role in the intrinsic and elevated resistance of *S.* Typhimurium to a wide range of compounds (51). Lama et al. (52) indicated that a nonsense mutation of *ksgA* caused resistance to amicoumacin A in methicillin-resistant *Staphylococcus aureus* (MRSA).

The region around *mutS* was thought to be an old region in the genome because it was part of the DNA mismatch repair system and SPI-1, which was acquired after *Salmonella* and *E. coli* separated from their common ancestor 100 million years ago (53). Diversities around conserved regions of the genome were identified and provided an insightful understanding of the evolutionary process. Region around *mutS* was hot spot for horizontal gene transfer and recombination events because this region was associated with pathogenicity and positive selection (54-56). For example, gene cluster 1 between *invH* and *mutS* included genes encoding ABC transport system protein. Recent studies reported that an ABC transporter gene was associated with the ability of *Salmonella* to acquire nutrients for survival during host infection (57) and drug resistance (58). Gene cluster 1 existed in lineage II and all outgroup genomes, but not in lineage III, suggesting the potential different pathogenic capability between two lineages. Moreover, the presence of transposable elements also could facilitate further genetic exchange in gene cluster 1. For example, gene cluster 3 in strain fish_Hong_Kong (TABLE II-5) was inserted at the 3' end of gene cluster 1, illustrating that the evolution of this region is an

ongoing process. Additionally, our data suggested that genes encoding restriction modification subunits of *Vibrio* were homologous to ORFs in gene cluster 3.

The limit of serological classification is that some unrelated strains were considered to be the same serotype (10, 59). As more data is available, distinct lineages of the same serotype are commonly found (32). *S.* Newport displays extensive genomic variation between lineages II and III, which are separated by other serotypes. Our data showed approximately 13,000 informative SNPs differences (SNPs that change once and define all members of these two major lineages) between these two lineages. Moreover, the pairwise distance matrix (TABLE II-2) suggested that the number of SNP differences between lineage III and any subgroup of lineage II was larger than that between lineage III and *S.* Hadar RI_05P066.

S. Newport has been proposed to be paraphyletic or polyphyletic (10-14) with distinct clonal lineages and it acted as a frequent donor or recipient of recombination events (47). According to the cross-link analysis, Sangal et al. (14) hypothesized that lineages II and III had arisen from a single lineage then differentiated or that recombination events frequently happened after lineages II and III shared a niche and then would merge in the future. However, our data indicated that the recombination events between lineages II and III were less frequent than those in lineage II. Lineages II and III were polyphyletic and were divergent from each other by other *Salmonella* serotypes. The remarkable inter-lineage distance (TABLE II-2) suggested that lineages II and III diverged early on in the serotype evolution of *S.* Newport and that they have evolved largely independently. Horizontal gene transfer and recombination events have been the major force for evolution of *S.* Newport (14) and our data supported that this pathogenic serotype has

extensive genomic diversity. It is likely that geographic and ecological structure provided physical proximity to facilitate the recombination events among bacteria, which may form subgroups of pathogen populations (60).

Additionally, a study of the pan-genome family tree indicated that *S.* Newport SL254 was separated from *S.* Newport SL317 in some single gene trees by another serotype, *S.* Hadar RI_05P066 (35), which showed close relatedness with lineage III in the present study (FIG II-1, TABLE II-2). This separation was confirmed by both the parsimony tree (FIG II-1) and phylogenetic dendrogram of *cas* genes (FIG II-5). In the current study, both of the MLST and *cas* genes could differentiate lineages II and III; however, they could not delineate strains within lineage II accurately. Therefore, full genome information or an improved MLST panel is needed to improve our understanding of the evolution of *Salmonella* (34, 47). Lineages II and III may have acquired the *cas* gene cluster from various sources. Although we do not fully understand the process of this genetic exchange, horizontal gene transfer also occasionally happened to housekeeping genes and this supports the hypothesis that the loci around *mutS* are hot spots for horizontal gene transfer.

Lineage II and the outgroup genomes possessed *ste* fimbrial operon between loci *relA* and *mazG* genes (TABLE II-6). The existence of the *ste* fimbrial operon may facilitate lineage II strains differing in their adhesion abilities and competing within various ecological environments (34, 61). den Bakker et al. (34) reported that genes enriched in different bacterial subpopulations could reveal various selective pressures acting on different subpopulations. Because genes between loci *relA* and *mazG* of these two lineages were both adjoining in *S.* Typhi CT18, this fimbrial operon may exist in lineage

III before it was lost. Thus, loci around *mutS* (62, 63) displayed mosaic structure because of recombination events.

References

1. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-625.
2. **Voetsch AC, Van Gilder TJ, Angulo FJ, Farley MM, Shallow S, Marcus R, Cieslak PR, Deneen VC, Tauxe RV.** 2004. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin Infect Dis **38 Suppl 3:**S127-134.
3. **CDC.** 2006. *Salmonella* Surveillance: Annual Summary. Atlanta, Georgia: US Department of Health and Human Services, CDC.
4. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.
5. **Patrick AD GF-X.** 2007. Antigenic formulae of the Salmonella serovars, 9th ed.
6. **Attorney S.** 2009. Multi-state *Salmonella* Newport outbreak leads to huge ground beef recall.
7. **Attorney S.** 2011. Sprout *Salmonella* Newport outbreak sickens six in Oregon and Washington.
8. **CDC.** 2010. Investigation update: multi-state outbreak of human *Salmonella* Newport infections linked to raw alfalfa sprouts. Atlanta, Georgia: US Department of Health and Human Services, CDC.
9. **Bell RL, Cao G, Meng J, Allard MW, Keys CE, Hill T, Ottesen AR, Brown EW.** 2012. *Salmonella* Newport Contamination of Produce: Ecological, Genetic,

and Epidemiological Aspects, p. In press, *Salmonella*: Classification, Genetics and Disease. Nova Publishers, Hauppauge NY.

10. **Beltran P, Musser JM, Helmuth R, Farmer JJ, 3rd, Frerichs WM, Wachsmuth IK, Ferris K, McWhorter AC, Wells JG, Cravioto A, et al.** 1988. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S.* choleraesuis, *S.* derby, *S.* dublin, *S.* enteritidis, *S.* heidelberg, *S.* infantis, *S.* newport, and *S.* typhimurium. Proc Natl Acad Sci U S A **85:**7753-7757.

11. **Harbottle H, White DG, McDermott PF, Walker RD, Zhao S.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J Clin Microbiol **44:**2449-2457.

12. **Sukhnanand S, Alcaine S, Warnick LD, Su WL, Hof J, Craver MP, McDonough P, Boor KJ, Wiedmann M.** 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella* enterica serotypes. J Clin Microbiol **43:**3688-3698.

13. **Torpdahl M, Skov MN, Sandvang D, Baggesen DL.** 2005. Genotypic characterization of *Salmonella* by multilocus sequence typing, pulsed-field gel electrophoresis and amplified fragment length polymorphism. J Microbiol Methods **63:**173-184.

14. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010.

Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

15. **Didelot X, Achtman M, Parkhill J, Thomson NR, Falush D.** 2007. A bimodal pattern of relatedness between the *Salmonella* Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination? Genome Res **17:**61-68.

16. **Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, Leclerc JE, Ravel J, Cebula TA.** 2011. Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. J Bacteriol **193:**3556-3568.

17. **Brown EW, Mammel MK, LeClerc JE, Cebula TA.** 2003. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens. Proc Natl Acad Sci U S A **100:**15676-15681.

18. **Mills DM, Bajaj V, Lee CA.** 1995. A 40 kb chromosomal fragment encoding *Salmonella* typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome. Mol Microbiol **15:**749-759.

19. **Hu Q, Coburn B, Deng W, Li Y, Shi X, Lan Q, Wang B, Coombes BK, Finlay BB.** 2008. *Salmonella enterica* serovar Senftenberg human clinical isolates lacking SPI-1. J Clin Microbiol **46:**1330-1336.

20. **Sorek R, Kunin V, Hugenholtz P.** 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol **6:**181-186.

21. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol **169:**5429-5433.

22. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

23. **Deveau H, Garneau JE, Moineau S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol **64:**475-493.

24. **Vale PF, Little TJ.** 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. Proc Biol Sci **277:**2097-2103.

25. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.

26. **McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK.** 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature **413:**852-856.

27. **Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S,**

O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848-852.

28. **Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK.** 2011. The origin of the Haitian cholera outbreak strain. N Engl J Med **364:**33-42.

29. **Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao G, Meng J, Stones R.** 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med **364:**981-982.

30. **Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD.** 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science **327:**469-474.

31. **Maki DG.** 2009. Coming to grips with foodborne infection--peanut butter, peppers, and nationwide *Salmonella* outbreaks. N Engl J Med **360:**949-953.

32. **Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW.** 2012. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **13:**32.

33. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

34. **den Bakker HC, Moreno Switt AI, Govoni G, Cummings CA, Ranieri ML, Degoricija L, Hoelzer K, Rodriguez-Rivera LD, Brown S, Bolchacova E, Furtado MR, Wiedmann M.** 2011. Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*. BMC Genomics **12:**425.

35. **Jacobsen A, Hendriksen RS, Aaresturp FM, Ussery DW, Friis C.** 2011. The *Salmonella* enterica pan-genome. Microb Ecol **62:**487-504.

36. **Klimke W, Agarwala R, Badretdin A, Chetvernin S, Ciufo S, Fedorov B, Kiryutin B, O'Neill K, Resch W, Resenchuk S, Schafer S, Tolstoy I, Tatusova T.** 2009. The National Center for Biotechnology Information's Protein Clusters Database. Nucleic Acids Res **37:**D216-223.

37. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.

38. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

39. **Goloboff P, Farris S, Nixon K.** 2008. TNT, a program for phylogenetic analysis. Cladistics **24:**774-786.

40. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221-224.

41. **Didelot X, Falush D.** 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics **175:**1251-1266.

42. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-2739.

43. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460-2461.

44. **Dunne EF, Fey PD, Kludt P, Reporter R, Mostashari F, Shillam P, Wicklund J, Miller C, Holland B, Stamey K, Barrett TJ, Rasheed JK, Tenover FC, Ribot EM, Angulo FJ.** 2000. Emergence of domestically acquired ceftriaxone-resistant *Salmonella* infections associated with AmpC beta-lactamase. JAMA **284:**3151-3156.

45. **Fey PD, Safranek TJ, Rupp ME, Dunne EF, Ribot E, Iwen PC, Bradford PA, Angulo FJ, Hinrichs SH.** 2000. Ceftriaxone-resistant *Salmonella* infection acquired by a child from cattle. N Engl J Med **342:**1242-1249.

46. **Park AK, Kim H, Jin HJ.** 2010. Phylogenetic analysis of rRNA methyltransferases, Erm and KsgA, as related to antibiotic resistance. FEMS Microbiol Lett **309:**151-162.

47. **Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D, Donnelly P.** 2011. Recombination and population structure in *Salmonella enterica*. PLoS Genet **7:**e1002191.

48. **Welch TJ, Fricke WF, McDermott PF, White DG, Rosso ML, Rasko DA, Mammel MK, Eppinger M, Rosovitz MJ, Wagner D, Rahalison L, Leclerc JE, Hinshaw JM, Lindler LE, Cebula TA, Carniel E, Ravel J.** 2007. Multiple antimicrobial resistance in plague: an emerging public health risk. PLoS One **2:**e309.

49. **Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, White DG, Cebula TA, Ravel J.** 2009. Comparative genomics of the IncA/C multidrug resistance plasmid family. J Bacteriol **191:**4750-4757.

50. **Matsumura K, Furukawa S, Ogihara H, Morinaga Y.** 2011. Roles of multidrug efflux pumps on the biofilm formation of *Escherichia coli* K-12. Biocontrol Sci **16:**69-72.

51. **Yamasaki S, Nagasawa S, Hayashi-Nishino M, Yamaguchi A, Nishino K.** 2011. AcrA dependency of the AcrD efflux pump in *Salmonella enterica* serovar Typhimurium. J Antibiot (Tokyo) **64:**433-437.

52. **Lama A, Pane-Farre J, Chon T, Wiersma AM, Sit CS, Vederas JC, Hecker M, Nakano MM.** 2012. Response of Methicillin-Resistant *Staphylococcus aureus* to Amicoumacin A. PLoS One **7:**e34037.

53. **Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, Petty NK, Kingsley RA, Baumler AJ, Nuccio SP, Contreras I, Santiviago CA, Maskell D, Barrow P, Humphrey T, Nastasi A, Roberts M, Frankel G, Parkhill J, Dougan G, Thomson NR.** 2011. *Salmonella bongori* provides insights into the evolution of the Salmonellae. PLoS Pathog **7:**e1002191.

54. **Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol **60:**1136-1151.

55. **Orsi RH, Sun Q, Wiedmann M.** 2008. Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*. BMC Evol Biol **8:**233.

56. **Lefebure T, Stanhope MJ.** 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol **8:**R71.

57. **Osborne SE, Tuinema BR, Mok MC, Lau PS, Bui NK, Tomljenovic-Berube AM, Vollmer W, Zhang K, Junop M, Coombes BK.** 2012. Characterization of DalS, an ATP-binding cassette transporter for D-alanine, and its role in pathogenesis in *Salmonella enterica*. J Biol Chem.

58. **Chang G.** 2003. Multidrug resistance ABC transporters. FEBS Lett **555:**102-105.

59. **Selander RK, Beltran P, Smith NH, Helmuth R, Rubin FA, Kopecko DJ, Ferris K, Tall BD, Cravioto A, Musser JM.** 1990. Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. Infect Immun **58:**2262-2275.

60. **Didelot X, Maiden MC.** 2010. Impact of recombination on bacterial evolution. Trends Microbiol **18:**315-322.

61. **Porwollik S (ed.).** 2011. *Salmonella*: from Genome to Function. Caister.

62. **Brown EW, Kotewicz ML, Cebula TA.** 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. Mol Phylogenet Evol **24:**102-120.

63. **Groisman EA, Ochman H.** 1997. How Salmonella became a pathogen. Trends Microbiol **5:**343-349.

TABLE II-1. General information of *Salmonella* Newport strains in this chapter.

Strain ID	Tree Label	PFGE Pattern Number	Antimicrobial Resistance Profile*	WGS Accession Number	Draft Genome Size (Mbp)	Number of Contigs
CVM 35185	bison_TN_2004	JJPX01.0218	SUL	AHTJ00000000	4.71	95
CVM 35199	caprine_TN_2004	JJPX01.0381	SUL	AHTK00000000	4.75	72
CVM 21539	chicken_MO	JJPX01.0030	NA	AHTL00000000	4.71	71
CVM 33953	ground_turkey_MD_2003	JJPX01.0502	NA	AHTM00000000	4.80	88
CVM 35188	equine_TN_2004_1	JJPX01.0025	SUL	AHTN00000000	4.71	66
CVM 21559	turkey_CO	NA	NA	AHTO00000000	4.74	64
CVM 19447	frog_Vietnam	JJPX01.3333	NA	AHTP00000000	4.67	59
CVM 19449	fish_Hong_Kong	JJPX01.0327	TET	AHTQ00000000	4.70	76
CVM 19567	fish_Vietnam	JJPX01.1947	NA	AHTR00000000	4.67	53
CVM 35202	equine_TN_2004_2	NA	SUL	AHTS00000000	4.96	72
CVM 21550	swine_TX	NA	NA	AHTT00000000	4.92	73
CVM 22513	cattle_NC_2003	JJPX01.0042	AMC,AMP,FOX,CHL,KAN,STR,SUL,TET,TIO	AHTU00000000	4.90	72
CVM 21538	chicken_GA	JJPX01.0238	NA	AHTV00000000	4.93	70
CVM 22425	cattle_AZ_2003	JJPX01.0014	AMC,AMP,FOX,CHL,STR,SUL,TET,TIO	AHTW00000000	4.93	69

CVM 22462	canine_AZ_2003	JJPX01.0014	AMC,AMP,FOX,CHL, STR,SUL,TET,TIO	AHTX00000000	5.02	384
CVM N18486	ground_turkey_NM_ 2008	JJPX01.0238	TET	AHTY00000000	4.93	85
CVM N1543	ground_beef_GA_20 04	JJPX01.0042	AMC,AMP,FOX,AXO, CHL,STR,SUL,TET,TIO	AHTZ00000000	4.89	77
CVM 21554	swine_IL_2001	NA	AMC,AMP,FOX,CHL, GEN,KAN,STR,SUL, TET,TIO	AHUA00000000	4.69	44
CVM 19443	shrimp_India	NA	NA	AHUB00000000	4.81	70
CVM 37978	spinach_CO_2008	JJPX01.0538	NA	AHUC00000000	4.80	49
CVM 19593	cheese_Mexico	JJPX01.0372	NA	AHUD00000000	4.65	74
CVM 19470	squid_Vietnam	NA	NA	AHUE00000000	4.73	84
CVM 19536	pepper_Vietnam	NA	NA	AHUF00000000	4.65	70
CVM 4176	pig_ear_CA	NA	NA	AHUG00000000	4.73	62
FDA 117	farm_1_VA_2007[#]	NA	NA	AJMN00000000	4.81	91
FDA 118	farm_15_VA_2007[#]	NA	NA	AJMO00000000	4.81	75
NA	*S*. Newport SL254	NA	AMP, CHL, GEN, STR,AXO,SUL,TET	ABEN01000000	4.83	0
NA	*S*. Newport SL317	NA	NA	ABEW00000000	4.95	63

*AMC = Amoxicillin/Clavulanic Acid, AMP = Ampicillin, FOX = Cefoxitin, AXO = Ceftriaxone, CHL = Chloramphenicol, GEN = Gentamicin, KAN = Kanamycin, STR = Streptomycin, SUL = Sulfamethoxazole or Sulfisoxazole , TET = Tetracycline, TIO = Ceftiofur. # These two samples were received from Eastern Shore of Virginia in 2007. Isolates may have been collected earlier than 2007.

TABLE II-2. Average pairwise distance (no. of nucleotide difference) for the major groups and outgroup genomes.

	subgroup III	Hadar	Typhimurium	subgroup IIC	subgroup IIA	subgroup IIB	Dublin	Gallinarum
Hadar	34418 (93)							
Typhimurium	36094 (90)	35900 (105)						
subgroup IIC	35048 (128)	37133 (147)	38640 (144)					
subgroup IIA	35627 (108)	37320 (152)	38893 (154)	17497 (95)				
subgroup IIB	36812 (106)	38529 (122)	38752 (131)	15605 (91)	25768 (85)			
Dublin	39879 (118)	40575 (133)	39275 (154)	40314 (175)	40878 (130)	41749 (136)		
Gallinarum	43027 (100)	43758 (159)	42824 (151)	43453 (158)	42666 (133)	44822 (124)	22070 (144)	
Kentucky	49260 (106)	49409 (80)	48612 (90)	50194 (96)	50694 (128)	48236 (146)	50955 (98)	53464 (96)

The valve refers to number of SNPs differences (standard deviation) between different selected groups and strains. The numbers of base differences per sequence from averaging over all sequence pairs between groups were shown.

TABLE II-3. Most variable genes that defining major lineages and subgroups

Gene	Locus A	Locus B	Nuc	AA	Position	# of SNPs	Group	Description
Genes define *S.* Newport Lineages II and III								
pduS	SNSL254_A2230	SNSL317_A3497	G->A	A/T	7	59		polyhedral body protein
thiC	SNSL254_A4496	SNSL317_A2323	C->G	G	72	59		thiamine biosynthesis protein
nrdA	SNSL254_A2462	SNSL317_A3257	C->T	I	186	51		ribonucleoside-diphosphate reductase
dmsA	SNSL254_A4651	SNSL317_A1619	T->C	R	315	51		anaerobic dimethyl sulfoxide reductase
ilvD	SNSL254_A4186	SNSL317_A2593	G->A	L	1275	47		dihydroxy-acid dehydratase
yacH	SNSL254_A0171	SNSL317_A1794	T->G	S	240	40		putative outer membrane protein
pduQ	SNSL254_A2229	SNSL317_A3498	G->T	A	111	38		propanediol utilization: propanol dehydrogenase
putA	SNSL254_A1218	SNSL317_A1473	G->A	A	873	38		trifunctional transcriptional regulator
rpoC	SNSL254_A4487	SNSL317_A2332	C->T	L	919	35		DNA-directed RNA polymerase subunit
gntR	SNSL254_A3810	SNSL317_A4927	T->C	S	255	34		putative GntR-family regulatory protein
hmt	SNSL254_A0021	SNSL317_A1949	C->T	T/G	1190	34		putative hydroxymethyltransferase
ytfN	SNSL254_A4775	SNSL317_A2173	C->T	A/V	59	33		putative periplasmic protein
norR	SNSL254_A3041	SNSL317_A4199	G->T	V	105	30		anaerobic nitric oxide reductase transcription regulator
yegQ	SNSL254_A2321	SNSL317_A3397	T->C	G	72	30		peptidase U32 family protein
tpiA	SNSL254_A4410	SNSL317_A2457	A->G	T/A	91	30		triosephosphate isomerase
araA	SNSL254_A0109	SNSL317_A1857	T->C	I	348	29		L-arabinose isomerase
hrpA	SNSL254_A1758	SNSL317_A0497	C->T	L	841	28		ATP-dependent helicase
hypF	SNSL254_A3044	SNSL317_A4196	C->T	F	264	28		hydrogenase maturation protein

infB	SNSL254_A3544	SNSL317_A3703	C->T	G	1026	27		translation initiation factor IF-2
sucA	SNSL254_A0795	SNSL317_A0018	A->G	E	843	27		2-oxoglutarate dehydrogenase E1 component
				Genes defines subgroup IIA				
recG	SNSL254_A4024	SEEN443_05060	C->G	L	858	36	IIA	ATP-dependent DNA helicase
parC	SNSL254_A3430	SEEN443_22651	G->C	S/T	170	30	IIA	DNA topoisomerase IV subunit A
nirC	SNSL254_A3748	SEEN443_16780	G->T	A/S	658	28	IIA	nitrite transporter
cysM	SNSL254_A2634	SEEN443_21787	G->A	A	480	25	IIA	cysteine synthase B
sgrR	SNSL254_A4141	SEEN443_14367	C->T	P	96	23	IIA	HTH-type transcriptional regulator
carB	SNSL254_A0071	SEEN443_13940	T->C	C	294	22	IIA	carbamoyl phosphate synthase large subunit
glmS	SNSL254_A4142	SEEN443_14372	C->T	R	222	20	IIA	glucosamine--fructose-6-phosphateaminotransferase
polB	SNSL254_A0103	SEEN443_13795	C->T	G	1080	20	IIA	DNA polymerase II
tktA	SNSL254_A3322	SEEN443_22146	T->G	P	294	19	IIA	transketolase
ligA	SNSL254_A4019	SEEN443_05035	A->G	T/A	103	19	IIA	NAD-dependent DNA ligase
pucJ	SNSL254_A3907	SEEN443_06684	A->G	Q	9	19	IIA	xanthine permease
cpdB	SNSL254_A4766	SEEN443_00595	C->T	A	780	18	IIA	3'-cyclic-nucleotide 2'-phosphodiesterase
	SNSL254_A3270	SEEN443_21941	G->A	A/T	55	18	IIA	putative inner membrane protein
levR	SNSL254_A4052	SEEN443_05190	G->C	Q/H	75	18	IIA	sigma-54 dependent transcription regulator
ybbP	SNSL254_A0562	SEEN443_07996	G->A	L	90	18	IIA	efflux ABC transporter permease protein
yggW	SNSL254_A3351	SEEN443_22286	C->T	D	183	17	IIA	putative oxidase
sgbU	SNSL254_A3954	SEEN443_06894	C->A	P/T	58	16	IIA	putative L-xylulose 5-phosphate 3-epimerase
phoU	SNSL254_A4134	SEEN443_14332	C->T	A	126	16	IIA	phosphate transport system regulatory protein
pepN	SNSL254_A1098	SEEN443_05597	C->A	P/Q	341	16	IIA	aminopeptidase N
metL	SNSL254_A4432	SEEN443_16720	G->A	T	816	16	IIA	bifunctional aspartate kinase II

			Genes define subgroup IIB					
hemL	SNSL254_A0223	SEEN447_13367	G->C	T	357	29	IIB	glutamate-1-semialdehyde aminotransferase
truB	SNSL254_A3542	SEEN447_18807	C->T	T	132	24	IIB	tRNA pseudouridine synthase B
gyrB	SNSL254_A4120	SEEN447_13058	C->T	Y	726	23	IIB	DNA gyrase subunit B
dho	SNSL254_A4791	SEEN447_14077	C->T	H/Y	199	22	IIB	dihydroorotase
purL	SNSL254_A2768	SEEN447_07165	C->T	L	175	22	IIB	phosphoribosylformylglycinamidine synthase
carA	SNSL254_A0070	SEEN447_20096	G->A	E/T	757	20	IIB	carbamoyl phosphate synthase small subunit
aceK	SNSL254_A4522	SEEN447_15307	C->T	S	741	19	IIB	bifunctional isocitrate dehydrogenase kinase
mgtA	SNSL254_A4804	SEEN447_14022	C->A	R	297	18	IIB	magnesium-translocating P-type ATPase
cyaA	SNSL254_A4221	SEEN447_11432	C->T	L	802	17	IIB	adenylate cyclase
uvrD	SNSL254_A4231	SEEN447_11487	C->T	L	1603	17	IIB	DNA-dependent helicase II
mtlA	SNSL254_A3963	SEEN447_08000	A->T	S	1222	16	IIB	pts system mannitol-specific eiicba component
creC	SNSL254_A4947	SEEN447_19716	T->C	L	199	14	IIB	sensory histidine kinase
glnS	SNSL254_A0745	SEEN447_07959	T->C	Y	576	14	IIB	glutaminyl-tRNA synthetase
trmE	SNSL254_A4127	SEEN447_13018	C->T	R	129	13	IIB	tRNA modification GTPase
mac	SNSL254_A4603	SEEN447_14967	T->C	F/S	65	13	IIB	putative integral membrane protein
yebU	SNSL254_A1989	SEEN447_10361	T->C	Y/H	7	13	IIB	paral putative rRNA methyltransferase
yniC	SNSL254_A1436	SEEN447_20726	G->A	S/N	203	12	IIB	phosphatase
treC	SNSL254_A4800	SEEN447_14037	G->A	V/I	1018	12	IIB	alpha phosphotrehalase
thiI	SNSL254_A0472	SEEN447_06040	C->T	S	1320	12	IIB	thiamine biosynthesis protein
barA	SNSL254_A3180	SEEN447_06891	C->T	N	1011	11	IIB	hybrid sensory histidine kinase
			Genes define subgroup IIC					
carB	SNSL317_A1896	SNSL254_A0071	T->C	V	132	36	IIC	carbamoyl phosphate synthase large subunit

yicL	SNSL317_A4701	SNSL254_A4029	A->G	I/V	100	26	IIC	alpha-xylosidase
ypfI	SNSL317_A1096	SNSL254_A2678	A->C	R	343	21	IIC	acetyltransferase
carA	SNSL317_A1897	SNSL254_A0070	A->T	I	60	18	IIC	carbamoyl phosphate synthase small subunit
yicH	SNSL317_A4702	SNSL254_A4028	C->A	P	453	17	IIC	AsmA family protein
speC	SNSL317_A3890	SNSL254_A3363	T->G	V/G	278	17	IIC	ornithine decarboxylase isozyme
guaA	SNSL317_A1071	SNSL254_A2703	T->C	L	346	15	IIC	GMP synthase
kdpD	SNSL317_A0049	SNSL254_A0764	T->C	L	1432	15	IIC	sensor protein KdpD
malP	SNSL317_A4954	SNSL254_A3788	G->A	V/M	1288	15	IIC	maltodextrin phosphorylase
malQ	SNSL317_A4955	SNSL254_A3787	G->A	A/T	646	15	IIC	4-alpha-glucanotransferase
torS	SNSL317_A4630	SNSL254_A4109	C->T	R/C	97	14	IIC	hybrid sensory histidine kinase
cysW	SNSL317_A1139	SNSL254_A2636	C->T	L	367	13	IIC	sulfate/thiosulfate transporter permease subunit
sgbU	SNSL317_A4778	SNSL254_A3954	C->G	Q/E	271	13	IIC	putative L-xylulose 5-phosphate 3-epimerase
dbpA	SNSL317_A0479	SNSL254_A1775	C->G	T/S	203	13	IIC	ATP-dependent RNA helicase
acrD	SNSL317_A1100	SNSL254_A2674	T->C	F	84	12	IIC	aminoglycoside/multidrug efflux system
nrfC	SNSL317_A1591	SNSL254_A4624	G->A	V/I	52	12	IIC	cytochrome c-type biogenesis protein
guaB	SNSL317_A1070	SNSL254_A2704	C->T	S	411	12	IIC	inosine-5'-monophosphate dehydrogenase
torA	SNSL317_A4634	SNSL254_A4105	T->C	G	60	12	IIC	trimethylamine-N-oxide reductase
kdpB	SNSL317_A0047	SNSL254_A0766	G->A	G/D	200	11	IIC	potassium-transporting ATPase subunit B
malZ	SNSL317_A2775	SNSL254_A0446	T->C	H	258	10	IIC	maltodextrin glucosidase
				Genes define node M				
rfaD	SNSL317_A4743	SNSL254_A3990	G->T	A/S	928	1	MDR	ADP-L-glycero-D-manno-heptose-6-epimerase
	SNSL317_A0618	SNSL254_A1638	G->A	V	222	1	MDR	glutaminase
	SNSL317_A2697	SNSL254_A0373	T->G	S/A	892	1	MDR	haloacetate dehalogenase H-1

fadH	SNSL317_A3774	SNSL254_A3480	T->C	C	1062	1	MDR	FAD/FMN-binding/pyridine nucleotide-disulphide oxidoreductase family protein
pstA	SNSL317_A4598	SNSL254_A4136	C->T	I	429	1	MDR	phosphate transporter permease subunit
yebZ	SNSL317_A0217	SNSL254_A2033	G->A	V/I	604	1	MDR	copper resistance protein D
ksgA	SNSL317_A1870	SNSL254_A0095	G->A	G/D	689	1	MDR	dimethyladenosine transferase
yedP	SNSL317_A0096	SNSL254_A2149	G->A	R/H	752	1	MDR	mannosyl-3-phosphoglycerate phosphatase
suhB	SNSL317_A1027	SNSL254_A2746	G->C	G/R	448	1	MDR	inositol monophosphatase
ydiY	SNSL317_A0820	SNSL254_A1441	A->T	S	165	1	MDR	outer membrane protein
hmt	SNSL317_A1949	SNSL254_A0021	T->C	L	2301	1	MDR	putative hydroxymethyltransferase
hycC	SNSL317_A4186	SNSL254_A3054	T->G	L/V	724	1	MDR	formate hydrogenlyase subunit 3
proQ	SNSL317_A0270	SNSL254_A1985	C->T	P/S	259	1	MDR	putative solute/DNA competence effector
uvrY	SNSL317_A0135	SNSL254_A2110	A->T	Q/L	284	1	MDR	response regulator
argO	SNSL317_A3941	SNSL254_A3305	A->G	A	317	1	MDR	arginine exporter protein
atpI	SNSL317_A4581	SNSL254_A4153	T->C	P	153	1	MDR	ATP synthase F0, I subunit
nuoG	SNSL317_A3209	SNSL254_A2507	C->T	T	231	1	MDR	NADH dehydrogenase subunit G
tppB	SNSL317_A0697	SNSL254_A1562	G->A	V	612	1	MDR	tripeptide transporter permease
yhjJ	SNSL317_A4850	SNSL254_A3885	G->A	A	870	1	MDR	
infB	SNSL317_A3703	SNSL254_A3544	C->T	L	307	1	MDR	translation initiation factor IF-2
ptsG	SNSL317_A4398	SNSL254_A1302	T->G	S/A	43	1	MDR	glucose-specific PTS system IIBC components
ulaA	SNSL317_A2201	SNSL254_A4744	C->T	T/I	83	1	MDR	ascorbate-specific PTS system enzyme IIC
ybfM	SNSL317_A0067	SNSL254_A0746	C->T	T	1167	1	MDR	outer membrane porin, OprD family
mac	SNSL317_A1571	SNSL254_A4603	C->T	Y	21	1	MDR	integral membrane protein
ygdH	SNSL317_A4058	SNSL254_A3192	C->T	C	555	1	MDR	lysine decarboxylase family protein
ynfM	SNSL317_A2693	SNSL254_A0369	T->G	F/V	286	1	MDR	permease
adiA	SNSL317_A1610	SNSL254_A4642	T->A	I	459	1	MDR	biodegradative arginine decarboxylase
wzzE	SNSL317_A2580	SNSL254_A4199	T->C	I/T	314	1	MDR	lipopolysaccharide biosynthesis protein

aroH	SNSL317_A0802	SNSL254_A1458	A->C	Q/P	449	1	MDR	phospho-2-dehydro-3-deoxyheptonate aldolase
speC	SNSL317_A3890	SNSL254_A3363	T->G	V/G	980	1	MDR	ornithine decarboxylase
yrbG	SNSL317_A3673	SNSL254_A3575	C->G	L/V	772	1	MDR	calcium/sodium:proton antiporter
fcl	SNSL317_A3437	SNSL254_A2290	G->A	V/M	94	1	MDR	GDP-L-fucose synthetase
cobD	SNSL317_A3518	SNSL254_A2210	C->A	G	555	1	MDR	cobalamin biosynthesis protein
				The 20 most variable genes				
carB	SNSL317_A1896	SNSL254_A0071				78		carbamoyl phosphate synthase large subunit
aceE	SNSL317_A1800	SNSL254_A0165				71		pyruvate dehydrogenase subunit E1
hrpA	SNSL317_A0497	SNSL254_A1758				69		ATP-dependent RNA helicase
putA	SNSL317_A1473	SNSL254_A1218				66		trifunctional transcriptional regulator
prpD	SNSL317_A2737	SNSL254_A0410				64		2-methylcitrate dehydratase
acnB	SNSL317_A1793	SNSL254_A0172				62		bifunctional aconitate hydratase 2
opdA	SNSL317_A4871	SNSL254_A3864				60		oligopeptidase A
thiC	SNSL317_A2323	SNSL254_A4496				59		thiamine biosynthesis protein
pduS	SNSL317_A3497	SNSL254_A2230				59		polyhedral body protein
dho	SNSL317_A2157	SNSL254_A4791				58		dihydroorotase
dmsA	SNSL317_A1619	SNSL254_A4651				55		anaerobic dimethyl sulfoxide reductase chain A
carA	SNSL317_A1897	SNSL254_A0070				54		carbamoyl phosphate synthase small subunit
thrA	SNSL317_A1968	SNSL254_A0002				54		bifunctional aspartokinase I/homoserine dehydrogenase I
yhiQ	SNSL317_A4872	SNSL254_A3863				53		methyltransferase
fadH	SNSL317_A3774	SNSL254_A3480				53		FAD/FMN-binding/pyridine nucleotide-disulfide oxidoreductase family protein
recG	SNSL317_A4706	SNSL254_A4024				53		ATP-dependent DNA helicase

yicJ	SNSL317_A2440	SNSL254_A4394	53	sugar (Glycoside-Pentoside-Hexuronide) transporter
yacH	SNSL317_A1794	SNSL254_A0171	52	putative outer membrane protein
nrdA	SNSL317_A3257	SNSL254_A2462	52	ribonucleotide-diphosphate reductase subunit alpha
ypfI	SNSL317_A1096	SNSL254_A2678	51	acetyltransferase

Variable genes were listed by their GenBank abbreviation and function description and by the locus to *S.* Newport SL254, SL317 and two Asian strains (strains from shrimp_India and frog_Vietnam). A representative nucleotide change observed within each gene is listed as well as whether this caused an amino acid change and to which phylogenetic group it was associated with from Figure 1. These genes and SNPs were the most valuable for the targeted resequencing and rapid subtyping methods for outbreak investigations. I listed 20 most variable genes that defining major and sub lineages. Moreover, we listed 33 informative SNPs that defining the MDR clade. # of SNPs means that the SNPs that changing once and defining members of major and subgroups.

TABLE II-4. Characteristics of genes/open reading frames (ORFs) between *invH* and *mutS* in gene cluster 1 of *S.* Newport SL254 and gene cluster 2 of strain chicken_MO.

ORF	Gene Name	Size (bps)	GC%	Best Blastp Hit				Super Family
				Description	Source	E Value	Locus Tag	
				Gene Cluster 1 in *S.* Newport SL254				
A3107		282	49.3	Putative ABC-type transport system	*S.* Typhi CT18	2e-45	NP_457295.1	DUF1778
A3108		528	47.2	Acetyltransferase, gnat family	*S.* Typhi CT18	7e-99	NP_457296.1	NA
A3109	*tnp*	438	55.9	Transposase	*Enterobacter cloacae*	1e-81	AAV66983.1	NA
A3110	*pphB*	657	41.7	Serine/threonine-specific protein phosphatase 2	*S.* Typhimurium LT2	1e-125	AAL21787.1	MPP
A3111		495	48.7	Membrane protein	*S.* Dublin CT_02021853	7e-92	ACH74700.1	NA
A3112		669	54.1	Hypothetical protein	*S.* Saintpaul SARA29	1e-125	EDZ12689.1	NA

ORF	Gene Name	Size (bps)	GC%	Best Blastp Hit				Super Family
				Description	Source	E Value	Locus Tag	
Gene Cluster 1 in *S.* Newport SL254								
11075	*insF*	738	52.3	Transposase InsF for insertion sequence IS3A/B/C/D/E/fA	*S.* Newport SL317	0	EDX51569.1	rve
11080	*insF*	171	55	Transposase InsF for insertion sequence IS3A/B/C/D/E/fA	*S.* Newport SL317	0	EDX51569.1	rve
11085		402	50.5	ISPsy11, transposase OrfA	*S.* Newport SL317	2e-92	EDX52090.1	NA
11090	*yis*	684	50.1	Integrase, catalytic region (ISPsy11, transposase OrfB)	*S.* Newport SL317	2e-170	EDX51974.1	rve
11095		258	47.3	ISEhe3 OrfA	*S.* Newport SL317	9e-57	EDX52144.1	HTH_Hin
11100	*tnpA*	90	46.7	Hypothetical protein	*S. Dublin* CT_02021853	6e-21	ACH75076.1	NA

Differences between Gene Cluster 1 and 2 demonstrated the mosaic genomic structure around *mutS* gene. Transposase and integrase were found in both sequences, indicating that both of them could be the hot spots for recombination events. The genes in both *S.* Newport SL254 and strain from chicken_MO are ordered top to bottom as their synteny on bacterial chromosome from 5' to 3'.

TABLE II-5. Characteristics of genes/open reading frames (ORFs) in gene cluster 3 of strain fish_Hong_Kong.

ORF	Gene Name	Size (bps)	GC%	Best Blastp Hit				Super Family
				Description	Source	E Value	Locus Tag	
18800	*fic*	618	49	FIC domain-containing protein (cell filamentation)	*Klebsiella pneumoniae* KCTC 2242	7e-129	AEJ99567.1	Fic
18795		171	45.6	hypothetical protein	*Klebsiella pneumoniae* KCTC 2242	1e-28	AEJ99568.1	NA
18790		810	44.4	hypothetical protein	*Acinetobacter baumannii* SDF	8e-160	CAP02803.1	HNHc
18785		2002	43.2	hypothetical protein	*Shewanella* sp. ANA-3	0	ABK49185.1	P-loop NTPase
18780		237	42.6	hypothetical protein	*Acinetobacter* baumannii SDF	7e-37	CAP02818.1	NA
18775	*hsdS*	1230	38.6	type I restriction enzyme specificity protein	*Vibrio splendidus* 12B01	0	EAP93672.1	Methylase_S
18770	*hsdM*	1536	41.3	type I restriction-modification system DNA-methyltransferase subunit M	*Vibrio metschnikovii* CIP 69.14	0	EEX37915.1	HsdM_N; AdoMet_MTases
18765		1473	45.8	hypothetical protein	*Yersinia intermedia* ATCC 29909	0	EEQ20934.1	NA

ORF	Gene Name	Size (bps)	GC%	Best Blastp Hit				Super Fam
				Description	Source	E Value	Locus Tag	
05488	*tnpR*	126	47.6	IS10 transposase	*S.* Kentucky	1e-22	ADK62113.1	NA
05493		453	45.3	hypothetical protein	*Yersinia intermedia* ATCC 29909	2e-108	EEQ20933.1	NA
05498		1692	44.3	hypothetical protein	*Yersinia intermedia* ATCC 29909	0	EEQ20932.1	DUF927
05503		879	41.3	hypothetical protein	*Yersinia intermedia* ATCC 29909	7e-137	EEQ20931.1	NA
05508		528	49.6	hypothetical protein	*Yersinia intermedia* ATCC 29909	6e-05	EEQ20929.1	NA
05513		558	35.3	acyltransferase	*Thermoanaerobacterium thermosaccharolyticum* DSM 571	6e-04	NC_014410.1	NA
05518	*int*	930	57.6	integrase/recombinase (Phage related)	*E. coli* WV_060327	0	EFW70420.1	DNA_BRE_C
05523	*radC*	453	53.9	putative phage DNA repair protein	*E. coli* SE15	6e-106	BAI54704.1	MPN

ORF	Gene Name	Size (bps)	GC%	Best Blastp Hit				Super Family
				Description	Source	E Value	Locus Tag	
05528		470	53.4	hypothetical protein	*E. coli* SE15	5e-111	BAI54703.1	NA
05533		423	42.6	hypothetical protein	*E. coli* SE15	8e-98	BAI54702.1	DUF2787
05538		489	50.5	hypothetical phage protein	*E. coli* SE15	8e-107	BAI54701.1	NA
05543	*nlp*	270	50.7	phage DNA-binding protein	*E. coli* SE15	7e-57	BAI54700.1	Nlp
05548		528	46.6	hypothetical protein	*E. coli* SE15	4e-127	BAI54699.1	NA
05553		360	49.7	hypothetical protein	*E. coli* SE15	7e-74	BAI54698.1	NA
05558		336	42.3	hypothetical protein	*E. coli* WV_060327	4e-72	EFW70428.1	NA
05563		1188	30.3	hypothetical protein	*Pseudomonas fluorescens* WH6	6e-54	EFQ66029.1	NA
05568	*int*	1287	47.1	Phage integrase	*Yersinia intermedia* ATCC 29909	0	EEQ20915.1	DNA_BRE_C

TABLE II-6. Characteristics of genes/open reading frames (ORFs) between *relA* and *mazG* of *S*. Newport SL254 and SL317.

ORF	Gene Name	Size (bps)	GC%	Best Blastp Hit				Super Family
				Description	Source	E Value	Locus Tag	
					S. Newport SL254			
A3171	*steA*	588	49.3	putative fimbrial subunit	*S*. Newport SL254	3e-136	ACF63661.1	Fimbrial
A3172	*steB*	2646	55.5	fimbrial usher protein	*S*. Newport SL254	0	ACF64468.1	PRK15223
A3173	*steC*	774	55	chaperone protein PapD	*S*. Newport SL254	0	ACF62389.1	Pili_assembly
A3174	*steD*	507	56.6	fimbrial subunit	*S*. Newport SL254	2e-118	ACF63171.1	Fimbrial
A3175	*steE*	471	50.7	fimbrial subunit	*S*. Newport SL254	1e-110	ACF62527.1	Fimbrial
A3176	*steF*	537	52.3	fimbrial subunit	*S*. Newport SL254	1e-128	ACF62131.1	Fimbrial
					S. Newport SL317			
A4073		288	48.3	putative addiction module antidote protein	*S*. Typhi CT18	2e-47	NP_457351.1	RHH_2
A4074		297	38.7	plasmid stabilization system protein, RelE/ParE family	*S*. Newport SL317	2e-51	ZP_02697812.2	Plasmid_stabil

I listed the detailed information of genes between *relA* and *mazG* genes. *S*. Newport SL254 and SL317 were selected. Our data indicated the genomic diversity of this region between Lineages II and III. Interestingly, ORF SNSL254_A3176 and SNSL317_A4073 were found adjoining together in *S*. Typhi CT18. The existence of *ste* fimbrial operon might enable Lineage II strains to infect variable hosts. The genes in both *S*. Newport SL254 and SL317 are ordered top to bottom as their synteny on bacterial chromosome from 5' to 3'.

FIG II-1. Parsimony phylogenetic tree of *S*. Newport and outgroup genomes.

This phylogenetic tree was reconstructed by TNT with 100,000 iterations based on 147,780 genome wide SNPs. All *S.* Newport strains were grouped into two major clusters, *S.* Newport Lineages II and III. Lineage II was further grouped into three subgroups, IIA, IIB and IIC. *S.* Newport displayed clear geographic structure. Asian strains were grouped together and divergent from ones from Americas. At the locus between *invH* and *mutS* genes, Lineage II and all outgroup genomes shared Gene Cluster 1; however, Lineage III strains shared Gene Cluster 2. Gene Cluster 3 was only found in strain from fish_Hong_Kong at the 3' end of Gene Cluster 1. GC1=Gene Cluster 1; GC2=Gene Cluster 2; GC3=Gene Cluster 3. Additionally, Node M includes most MDR strains in the current study.

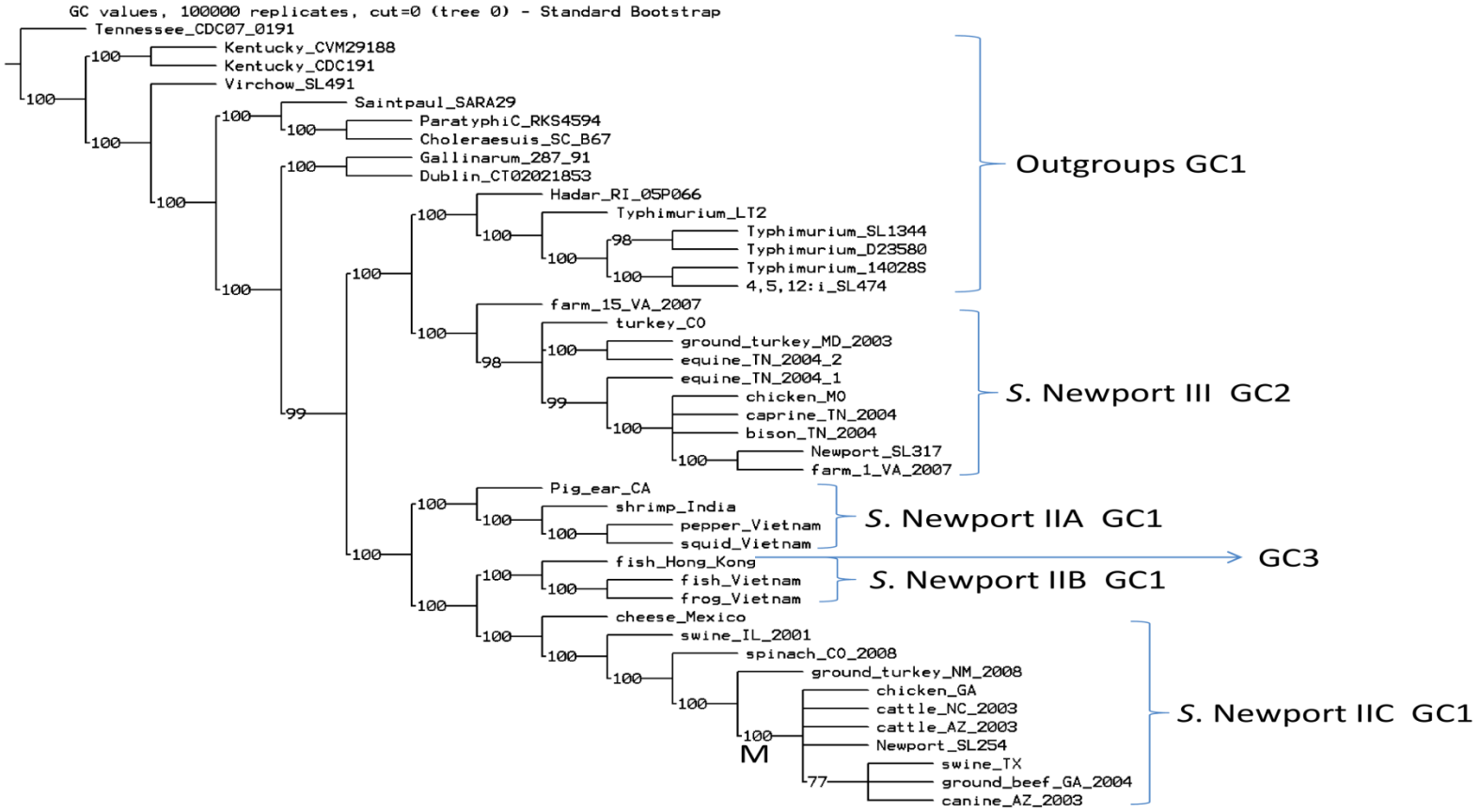

GC values, 100000 replicates, cut=0 (tree 0) - Standard Bootstrap
Tennessee_CDC07_0191
Kentucky_CVM29188
Kentucky_CDC191
Virchow_SL491
Saintpaul_SARA29
ParatyphiC_RKS4594
Choleraesuis_SC_B67
Gallinarum_287_91
Dublin_CT02021853
Hadar_RI_05P066
Typhimurium_LT2
Typhimurium_SL1344
Typhimurium_D23580
Typhimurium_14028S
4,5,12:i_SL474
farm_15_VA_2007
turkey_CO
ground_turkey_MD_2003
equine_TN_2004_2
equine_TN_2004_1
chicken_MO
caprine_TN_2004
bison_TN_2004
Newport_SL317
farm_1_VA_2007
Pig_ear_CA
shrimp_India
pepper_Vietnam
squid_Vietnam
fish_Hong_Kong
fish_Vietnam
frog_Vietnam
cheese_Mexico
swine_IL_2001
spinach_CO_2008
ground_turkey_NM_2008
chicken_GA
cattle_NC_2003
cattle_AZ_2003
Newport_SL254
swine_TX
ground_beef_GA_2004
canine_AZ_2003
M
Outgroups GC1
S. Newport III GC2
S. Newport IIA GC1
GC3
S. Newport IIB GC1
S. Newport IIC GC1

FIG II-2. MLST analysis of *S*. Newport and outgroup genomes.

Seven housekeeping genes were selected and MLST dendrogram was performed by TNT with 100,000 iterations. *S*. Newport was divided into two major clusters, which were separated by outgroup genomes. Lineage II was divided into three subgroups, which display minor differences compared with the parsimony tree. Subgroup IIA showed closer relatedness with IIC than IIB.

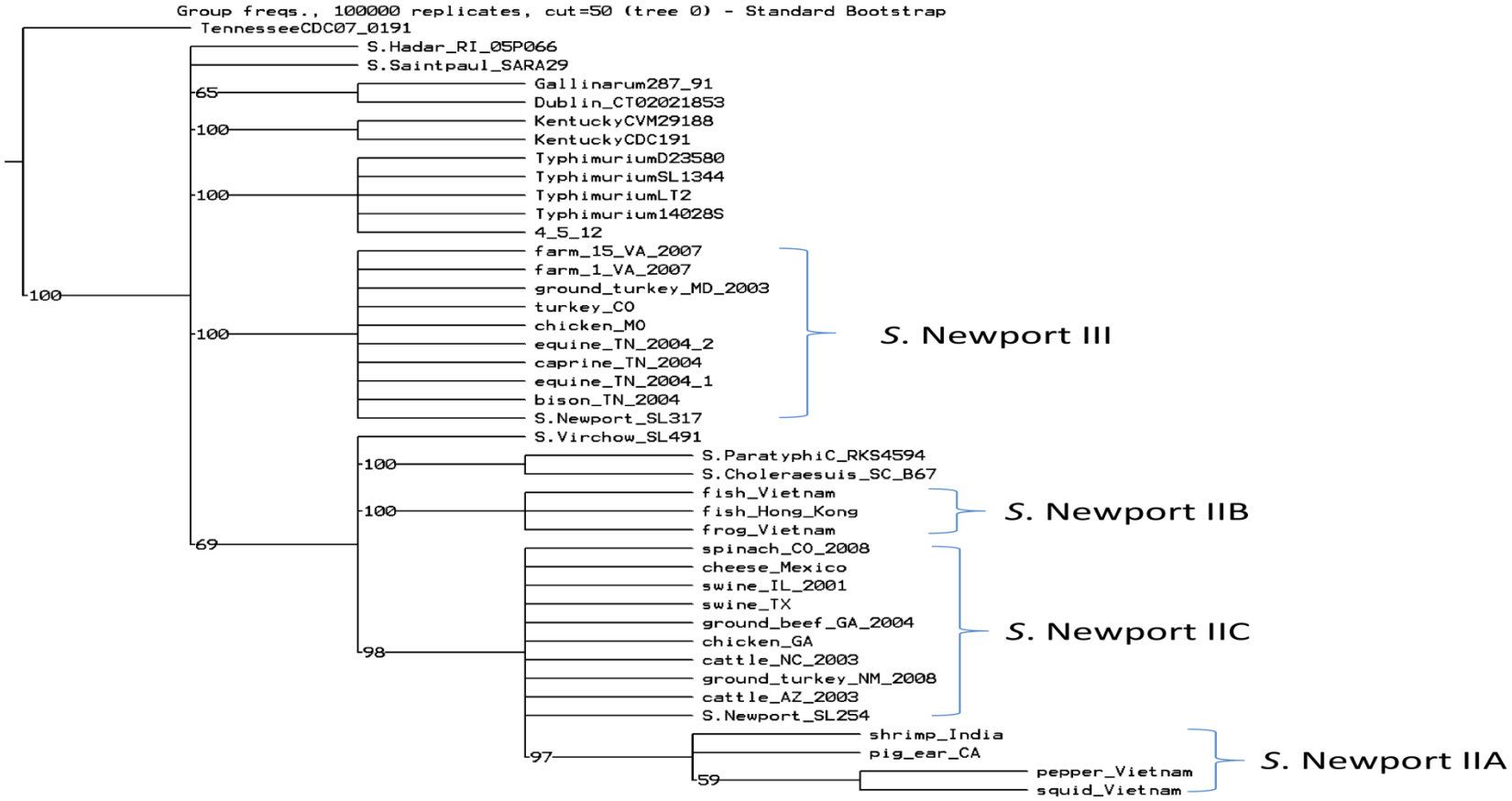

Group freqs., 100000 replicates, cut=50 (tree 0) - Standard Bootstrap
TennesseeCDC07_0191
S.Hadar_RI_05P066
S.Saintpaul_SARA29
Gallinarum287_91
Dublin_CT02021853
KentuckyCVM29188
KentuckyCDC191
TyphimuriumD23580
TyphimuriumSL1344
TyphimuriumLT2
Typhimurium14028S
4_5_12
farm_15_VA_2007
farm_1_VA_2007
ground_turkey_MD_2003
turkey_CO
chicken_MO
equine_TN_2004_2
caprine_TN_2004
equine_TN_2004_1
bison_TN_2004
S.Newport_SL317
S. Newport III
S.Virchow_SL491
S.ParatyphiC_RKS4594
S.Choleraesuis_SC_B67
fish_Vietnam
fish_Hong_Kong
frog_Vietnam
S. Newport IIB
spinach_CO_2008
cheese_Mexico
swine_IL_2001
swine_TX
ground_beef_GA_2004
chicken_GA
cattle_NC_2003
ground_turkey_NM_2008
cattle_AZ_2003
S.Newport_SL254
S. Newport IIC
shrimp_India
pig_ear_CA
pepper_Vietnam
squid_Vietnam
S. Newport IIA
100
65
100
100
100
100
69
100
98
97
59

FIG II-3. Genomic organization comparisons between subgroups.

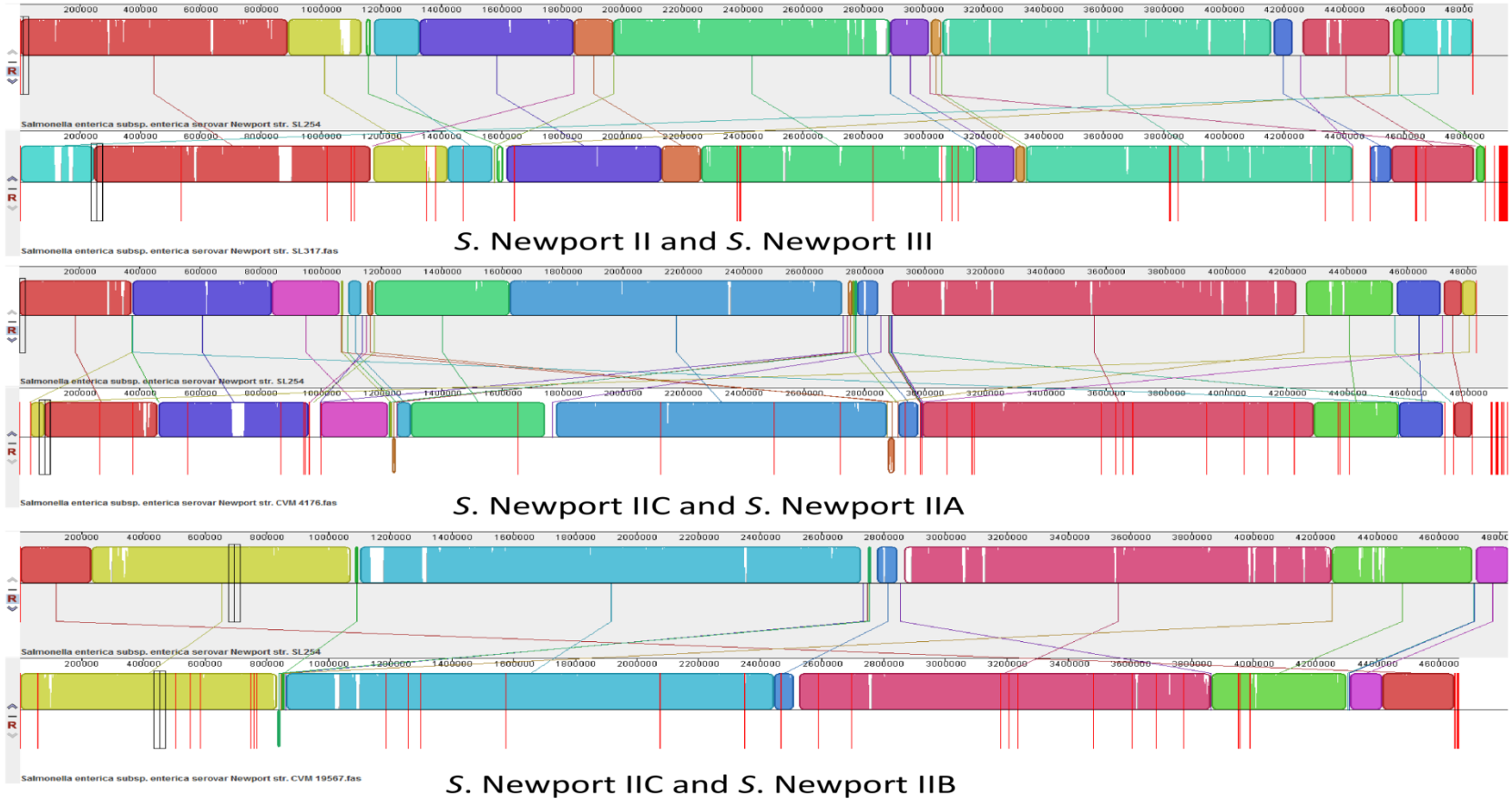


FIG II-4. ClonalFrame analyses of recombination events.

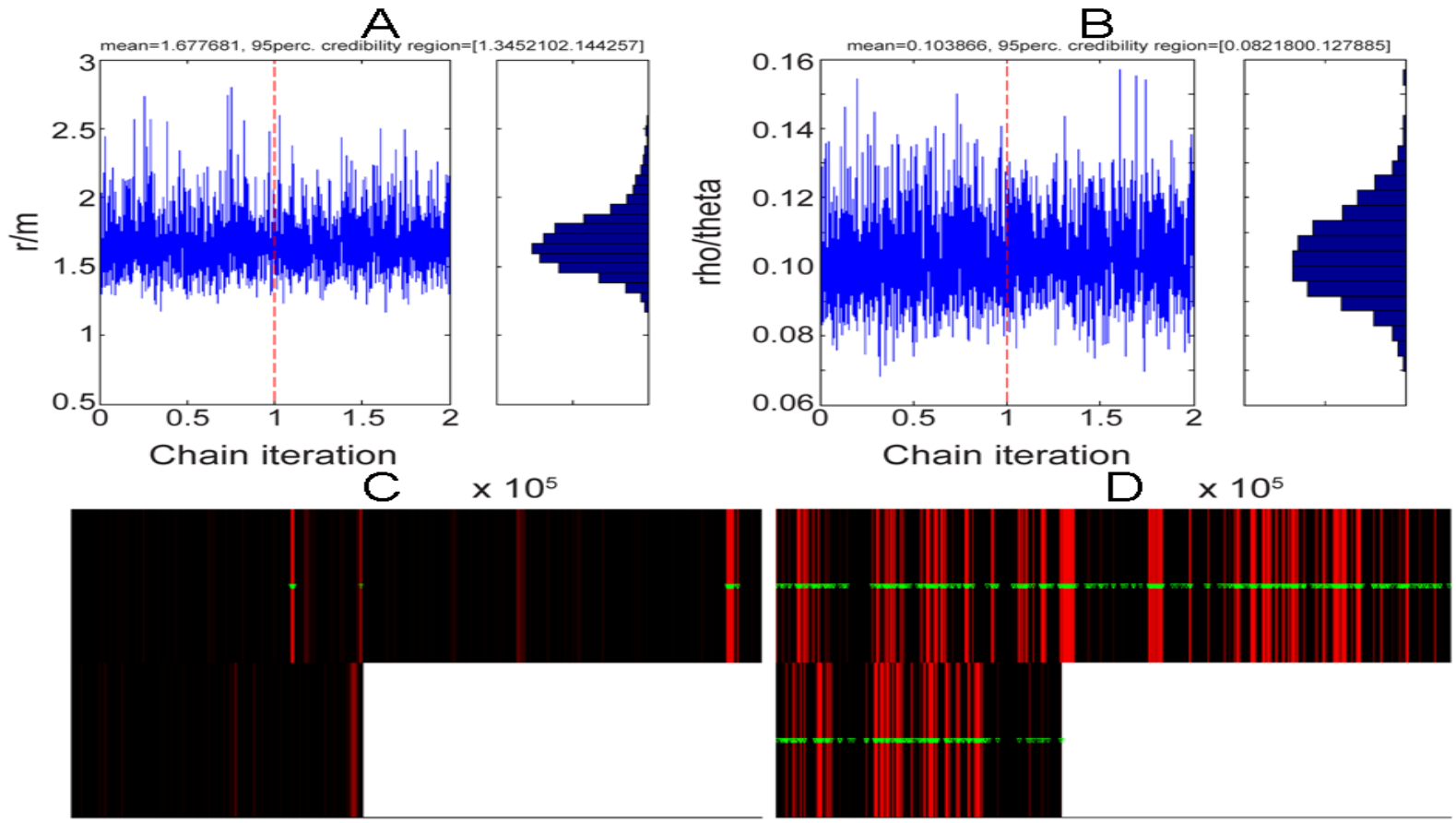


FIG II-5. Parsimony phylogenetic tree for *cas* genes.

I constructed this parsimony tree with 100,000 iterations by TNT based on concatenated sequences of the *cas* genes. This dendrogram indicated that *cas* genes of Lineages II and III were originated from distinct sources.

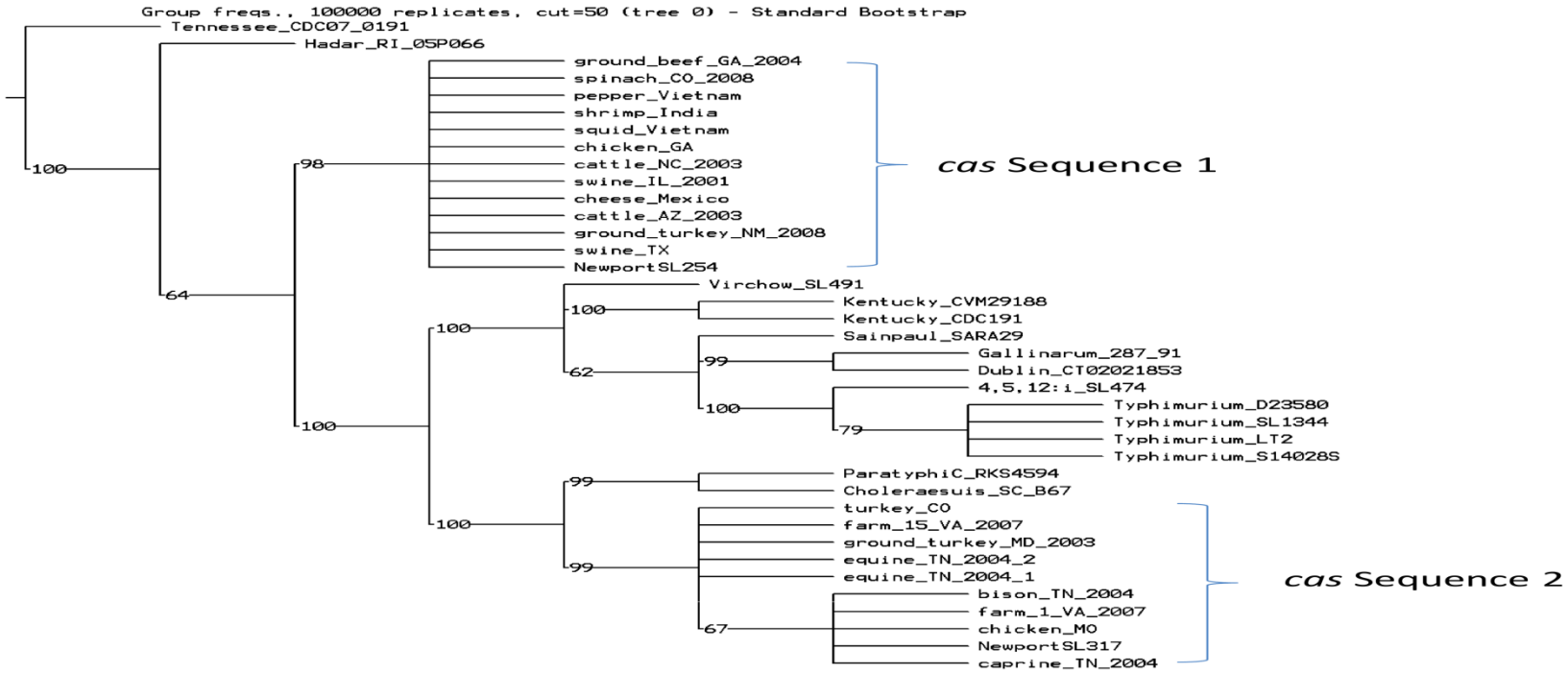
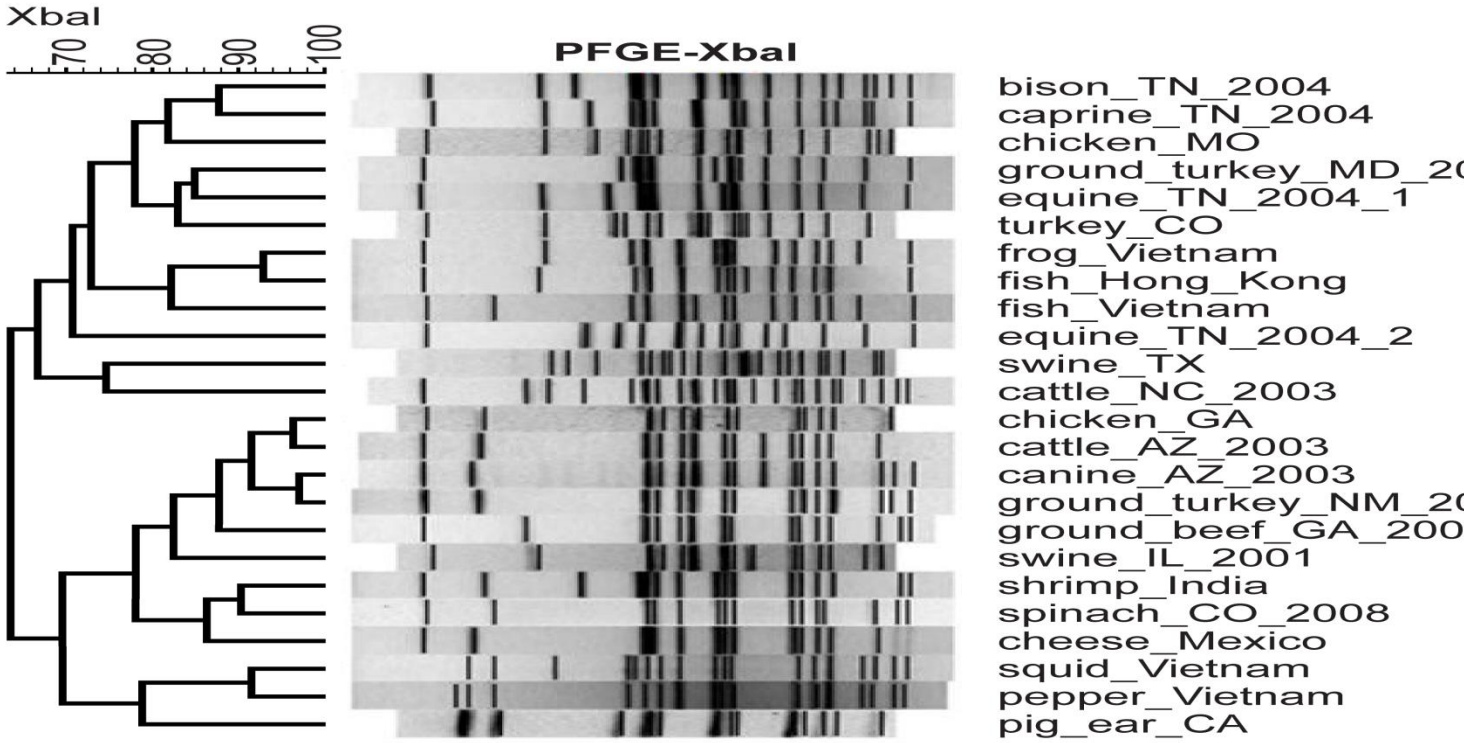


FIG II-6. Pulsed Field Gel Electrophoresis (PFGE) profile digested with *Xba*I.

I performed PFGE analysis of 24 *S.* Newport strains (without two environmental farm isolates) isolated from diverse sources and geographic locations. PFGE profiles divided these strains into two major clusters with different groupings compared with the phylogenetic tree based on whole genome wide SNPs.



CHAPTER III: GENETIC DIVERSITY OF *SALMONELLA* PATHOGENICITY ISLANDS SPI-5 AND SPI-6 IN *SALMONELLA* NEWPORT

Abstract

This chapter was submitted to Microbiology. *Salmonella* Newport is one of common serovars causing foodborne salmonellosis outbreaks in the United States. It consists of three lineages with extensive genetic diversity. Most *S.* Newport strains from North America belong to *S.* Newport lineages II and III. A total of 28 strains of lineages II and III from diverse sources and geographic locations were analyzed using whole genome sequencing technology. Because of the importance of *Salmonella* pathogenicity islands 5 and 6 (SPI-5 and SPI-6) in virulence activity of pathogenic *Salmonella*, the presence and genetic diversities of these two SPIs may be highly associated with *S.* Newport pathogenicity. SPI-5, encoding translocated effector proteins of SPI-1 and SPI-2, was identified in all *S.* Newport strains with variations. It contained two genomic islands (SPI5-GI1 and SPI5-GI2) of over 40 kb encoding bacteriophage genes between tRNA-ser and *pipA* and 146 single nucleotide polymorphisms (SNPs). SPI5-GI1 was identified in all *S.* Newport multi-drug resistant strains. There were 39 lineage-defining SNPs identified including 18 none-synonymous SNPs. SPI-6 was also present in all *S.* Newport strains except three Asian strains in subgroup IIA. The Asian strains shared a common genomic island at the same locus of SPI-6. The *saf* fimbrial operon downstream of SPI-6 was present in the *S.* Newport strains. The phylogenetic trees of SPI-6 constructed with 937 SNPs showed that all *S.* Newport showed clear geographic structure at the lineage

level. These findings illustrated the genetic diversity of these important SPIs and implied the potential differences of virulence in these *S.* Newport strains.

Introduction

Non-typhoidal *Salmonella* cause 1.4 million foodborne illness cases annually in the United States, accounting for 11% of all foodborne infections (1). *Salmonella enterica* subspecies *enterica* serovar Newport (*S*. Newport) is one significant serovar associated with foodborne outbreaks and causes over 100,000 infections each year in the United States (2, 3). It has been responsible for several multistate foodborne outbreaks associated with ground beef and tomatoes since 2002 (3-5). *S.* Newport is polyphyletic with extensive genetic diversity and consists of three lineages based on multilocus sequence typing (MLST) analysis (6). Most *S.* Newport strains from Europe belong to lineage I whereas most strains from North America belong to lineages II and III (6). I performed whole genome sequencing analysis of 28 *S.* Newport strains from diverse sources and geographic locations, and grouped them into lineages II and III with clear geographic structure (7). The strains from Asia were decoupled from those from the Americas. *S.* Newport strains from Asia belonged to lineage II (7). No *S.* Newport lineage I strains have been sequenced.

Pathogenicity islands are blocks of genes encoding various virulent determinants and usually absent in non-pathogenic strains of the same or closely related species (8). There have been 22 *Salmonella* pathogenicity islands (SPIs) identified (8). The selected *S.* Newport genomes contained conserved contents in SPI-1 through SPI-4 (unpublished data) and showed extensive diversities at the region around *mutS* downstream of SPI-1 (7).

SPI-5 was first identified in the *S.* Dublin genome between tRNA-*ser*T and *copR* consisting of five genes (*pipA*, *pipB*, *pipC*, *sopB* and *pipD*) (9). These five genes

displayed a high similarity with genes of bacteriophages *Gifsy*-1 and *Gifsy*-2, such as *pipA* (10). SPI-5 plays an important role in pathogenicity, such as encoding effectors of SPI-1 and SPI-2 (8). *sopB* encoded a translocated effector protein of type III secretion systems (T3SS) in SPI-1 under control of *hilA*, whereas *pipB* encoded translocated effector of T3SS in SPI-2 under control of *ssrAB* (11, 12). Additionally, SPI-5 is thought to contribute to the colonization of the spleen in chickens (13). Mutations among SPI-5 genes significantly reduced the enteropathogenicity of *Salmonella* (9).

SPI-6 was located between tRNA-*asp*V and *sinR* at centisome 7 in *Salmonella* genomes containing the type six secretion system (T6SS) and the *Salmonella* atypical fimbriae (*saf*) fimbrial gene cluster (8). SPI-6 possesses different contents in various serovars. For example, it has a 47 kb island in *S*. Typhimurium (14) and a 59 kb island in *S*. Typhi (15). T6SS is widespread in bacteria (16) and possesses diverse functions (17, 18), such as the ability to encode virulence factors (19) and to mediate antagonistic interactions between bacteria (20). Folkesson (21) reported that the deletion of SPI-6 reduced the invasion activity of *S*. Typhimurium into Hep2 cells. Some *Salmonella* serovars such as *S*. Dublin and *S*. Weltevreden contained two T6SS-encoding SPIs (SPI-6 and SPI-19), whereas *S*. Gallinarum, *S*. Enteritidis and *S*. Agona only carried SPI-19. However, no T6SS was identified in certain genomes including *S*. Virchow, *S*. Paratyphi B and *S*. Javiana (17). The *saf* gene clusters are located downstream of SPI-6 and are present in most clinical isolates of *S. enterica* subsp. *enterica* (14, 22). The *saf* operon does not contribute to virulence in mouse encoding non-fimbrial adhesion elements (14).

The object of the current chapter is to investigate the genetic diversity of SPI-5 and SPI-6 in *S.* Newport lineages II and III and to identify potential markers in these important islands for further subtyping and detection.

Materials and Methods

Genomes

Twenty-eight *S.* Newport strains from diverse sources and geographic locations (7, 23) and 11 outgroup genomes (7, 24, 25) were used in the current chapter, including *S.* Tennessee CDC07_0191 (ACBF00000000), *S.* Kentucky CVM29188 (ABAK00000000), *S.* Kentucky CDC191 (ABEI00000000), *S.* Gallinarum 287/91 (AM933173.1), *S.* Dublin CT02021853 (CP001144.1), *S.* Hadar RI_05P066 (ABFG00000000), *S.* Typhimurium LT2 (NC_003197.1), *S.* Typhimurium SL1344 (NC_016810.1), *S.* Typhimurium D23580 (NC_016854.1), *S.* Typhimurium 14028S (CP001363.1) and *S.* 4,[5],12:i:- SL474 (ABAO00000000).

Phylogenetic tree construction

Whole genome parsimony tree was reconstructed based on 131,855 informative SNPs using tree analysis using new technology (TNT) (26) with finding minimum tree length 20 times and 100,000 bootstrapping iterations. Moreover, we performed multiple sequence alignment using MULCLE (27) and identified 146 SNPs in SPI-5, 937 SNPs in SPI-6 (excluding *saf* genes) and 355 SNPs in *saf* genes. Parsimony trees of SPI-5, SPI-6, and *saf* genes were reconstructed using TNT and the same parameters as above.

Genetic characterizations of SPI-5 genomic islands 1 and 2 (SPI5-GI1 and SPI5-GI2), and SPI-6 genomic island 1 (SPI6-GI1)

Genetic organizations between the SPI5-GIs and between SPI-6 and SPI6-GI1 were determined using Mauve (28). The best match of genes in SPI5-GI1, SPI5-GI2, and SPI6-GI1 was performed using blastp (29), followed by verification using tblastn (29).

Distance matrix

MEGA 5.05 (30) was used to calculate evolutionary distances (number of differences) over sequence pairs with 10,000 bootstrap iterations for SPI-5, SPI-6, and *saf* genes.

Results

Phylogenetic tree of genome data

A whole genome phylogenetic tree was constructed with more than 130,000 SNPs (FIG III-1). To better display the evolutionary relationship between *S.* Newport strains, a total of 11 genomes were selected as outgroups. There were six equally most parsimonious trees with the same branch order at subgroup level, meaning that the distributions of taxa in each *S.* Newport subgroup were same in all resulting trees. *S.* Newport showed identical phylogenies as our previous report (7), in which 15 outgroups were chosen.

Genetic diversity of *Salmonella* pathogenicity island 5

SPI-5 was present in the 28 *S.* Newport and 11 outgroup genomes. SPI-5 variations included indels and mutations. Genomic islands encoding bacteriophage genes were identified between tRNA-*ser* and *pipA* in certain genomes, which were designated as SPI-5 genomic islands 1 and 2 (SPI5-GI1 and SPI5-GI2) (FIG III-1 and III-2). SPI5-GI1 was present in node M strains, which contained all multi-drug resistant (MDR) strains in the current study. SPI5-GI2 was only found in strain shrimp_India. The SPI5-GIs were both more than 40 kb in length containing prophage genes (TABLE III-1 and III-2).

A gene cluster in SPI5-GI1 (SNSL254_A1155 to SNSL254_A1177, 5' to 3') encoding bacteriophage genes was present in several *Salmonella* serovars, including *S.* Typhi CT18, *S.* Paratyphi B SPB7, *S.* Paratyphi C RKS4594, and *S.* Choleraesuis SC-B67. These four genomes did not contain insertions in SPI-5. SPI5-GI2 possessed genes showing high identities with those in *S.* Weltevreden (TABLE III-2). A gene cluster in SPI5-GI2 (SEEN443_12678 to SEEN443_12753, 5' to 3') was identified in *S.*

Weltevreden HI_N05-537, *S.* Newport SNSL317, *S.* Typhimurium DT104, and *S.* Saintpaul SARA29. These various serovars genomes did not contain insertions in SPI-5. According to the blast matches, there were 74% genes in SPI5-GI1 and 52% genes in SPI5-GI2 annotated as hypothetical proteins or bacteriophage genes. Based on current annotation, there are no genes related to virulence or antimicrobial resistance present. However, both SPI5-GIs harbored gene encoding lytic transglycosylase, which is a common bacterial enzyme acting on peptidoglycan (31).

SPI-5 possessed 146 SNPs in five genes, which were used to construct the phylogenetic tree (FIG III-3). There were 227 equally most parsimonious trees with the same branch order at the lineage level showing that *S.* Newport in each lineage were clustered together but separated by outgroups in all resulting trees. Similar to the whole genome tree, the SPI-5 phylogenetic tree (FIG III-3) showed lineages II and III were separated by outgroup genomes. There were two differences between the genome tree and the SPI-5 tree. First, for the SPI-5 tree, each lineage was conserved and the SNPs could not distinguish *S.* Newport at subgroup level in lineage II. Second, in the SPI-5 tree, lineage II showed a close relationship with *S.* Typhimurium and *S.* Hadar, but not *S.* Newport lineage III. In addition, the trees of the five individual genes in SPI-5 demonstrated the same phylogenies (data not shown).

Pairwise distance matrix of SPI-5 revealed evolutionary divergence between *S.* Newport and outgroup genomes (TABLE III-3). The average differences between lineages II and III were 40 SNPs but only 18 SNPs between *S.* Typhimurium and lineage II. *S.* Dublin CT_02021853 was more closely related to lineage III (18 SNPs) than to other groups. *S.*

Gallinarum 287/91 also separated lineages II and III, with a closer phylogenetic relationship with lineage II.

A total of 39 SNPs in SPI-5 defined lineages II and III, meaning that all strains in each lineage shared the same nucleotides (four SNPs in *pipA*, nine in *pipB*, seven in *pipC*, five in *sopB* and 14 in *pipD*) (TABLE III-4). There were 18 of 39 SNPs leading to non-synonymous substitutions, including 4, 7, 2, 2 and 3 non-synonymous substitutions in *pipA*, *pipB*, *pipC*, *sopB* and *pipD*, respectively. Thus, 100% and 78% of these lineage-defining SNPs in *pipA* and *pipB* lead to amino acid changes (TABLE III-4).

Genetic diversity of *Salmonella* pathogenicity island 6

SPI-6 was present in all *S.* Newport genomes except the Asian strains in subgroup IIA including shrimp_India, squid_Vietnam, and pepper_Vietnam (FIG III-1). These three Asian strains possessed a common gene cluster at the same location as SPI-6, which was termed SPI-6 genomic island 1 (SPI6-GI1). SPI6-GI1 also was identified in *S.* Virchow SL491, which did not contain SPI-6. The genetic organizations of SPI-6 and SPI6-GI1 displayed extensive differences (FIG III-4).

General characterizations of genes in SPI6-GI1 were listed (TABLE III-5). The best blastp matches for SPI6-GI1 included *S.* Paratyphi B SPB7 and *S.* Gallinarum 287/91. There were 9 and 13 genes in SPI6-GI11 identified in *S.* Paratyphi B SPB7 and *S.* Gallinarum 287/91, respectively. These genes were present between tRNA-*asp* and *sinR* in *S.* Paratyphi B SPB7 and *S.* Gallinarum 287/91. According to the annotation of blast matches, SPI6-GI1 did not carry any known genes related to virulence or antimicrobial resistance.

SPI-6 contained 937 variable SNPs (excluding the *saf* genes), which were used to construct a phylogenetic tree (FIG III-5). There were 208 equally parsimonious trees with the same branch orders at the subgroup level. Compared to the genome phylogenetic tree, SPI-6 tree showed clear geographic structure at lineage level, meaning that three Asian strains in subgroup IIB was decoupled with all American strains belonging to lineages II and III. In the American strains, lineage III and subgroup IIC were separated. There was 672 SNPs between IIB and IIC, whereas only 222 SNPs between IIB and *S*. Hadar RI_05P066.

All 28 *S*. Newport strains contained *safABCD* downstream of SPI-6. *safA* was not present in two *S*. Kentucky strains, *S*. Hadar RI_05P066, *S*. Dublin CT_02021853, and *S*. 4,[5],12:i:- SL474. Pseudogenes were identified in the *saf* cluster, such as *safB* in *S*. Kentucky CVM29188 (SeKA_A4625), *S*. Dublin CT_02021853 (SeD_A0329) and *safC* in *S*. Gallinarum 287/91 (SG0310).

Similar to the SPI-6 gene tree, the *saf* tree showed the six Asian strains were grouped together and were separated from all the American strains by *S*. Typhimurium group (FIG III-6). Subgroup IIB strains clustered together. Strain shrimp_India displayed distant relationship with the other five Asian strains. *S*. Tennessee contained 71 SNPs differences with IIA&B and contained only 21 SNPs with shrimp_India (FIG III-6, TABLE III-3). In the American group, lineage III and subgroup IIC were separated by *S*. Gallinarum 287/91. Strain pig_ear_CA in IIA seemed to be an exception, showing close relationship with lineage III. Additionally, one gene cluster consisting of *tcfABCD* fimbrial operon, *tinR*, and *tioA* was only present in strains squid_Vietnam and pepper_Vietnam subgroup IIA (FIG III-4).

Discussion

In the present study, we investigated the genetic diversity of SPI-5 and SPI-6 in *S.* Newport lineages II and III. The *S.* Newport genomes all contained SPI-5, which had two SPI5-GIs and 146 SNPs. SPI-6 was also present in all *S.* Newport genomes except the three Asian strains in subgroup IIA, which possessed SPI6-GI1 at the same location of SPI-6. Moreover, the presence of *saf* genes downstream of SPI-6 could distinguish lineage II subgroups and strains from Asia and the Americas.

SPIs play significant roles in causing human illness (8). Because of the similarities of DNA sequence between bacteriophages and pathogenicity islands (PAIs), PAIs likely have originated from phage via horizontal gene transfer (HGT), such as SPIs (8), *Vibrio* pathogenicity island (VPI) and *Staphylococcus aureus* pathogenicity island 1 (SaPI1) (32). Knodler (11) demonstrated that SPI-5 genes may be acquired through HGT from lambdoid phages, including *Gifsy*-1 and *Gifsy*-2. Bacteriophages may play key roles in virulence activities of *Salmonella*, facilitating bacteria to survive in different ecological environments. For example, bacteriophage has been important for the genomic evolution of *S.* Montevideo and *S.* Enteritidis (33, 34).

SPI5-GIs containing bacteriophage genes also may play significant roles in virulence. I hypothesized that SPI5-GI1 was originally introduced in the most recent common ancestor of node M via HGT and transmitted it vertically to the offspring strains. SPI5-GI1 may become functionally compatible to the genomes in node M. The presence of SPI5-GI2 indicated that the region between tRNA-*ser* and *pipA* might be a hot spot for independent acquisitions of genetic elements. Since the function of genes in both SPI5-GIs were not well annotated, functional studies of SPI-5 with and without these GIs

might be important to undertake. Both SPI5-GIs contained genes encoding lytic transglycosylases, which might be used as target for broad-spectrum antibiotics (31). The SPI-5 genes could be the targets for resequencing and biomarkers to differentiate lineages II and III in rapid detection. In the lineage-defining SNPs, all mutations in *pipA* and 78% of the mutations in *pipB* lead to non-synonymous substitutions indicating that these two genes were under positive selections. Positive selection played critical roles in the evolution of different bacterial pathogens, as it accounts for 1.2% of *Salmonella* core genome including virulence genes (35). In one study focusing on positive selection in pathogenic *Salmonella*, three genes showed evidence of positive selection among SPI-1 through SPI-6, including *pipB* (SPI-5) and *safC* (SPI-6) (35). Since *S.* Newport was not included, *pipA* may show serotype-specific positive selection in *S.* Newport. Non-synonymous substitutions in SPI-5 may have influenced the pathogenicity of the genomes. Two non-synonymous substitutions were identified in domain CHASE3 in *pipA*, which was associated with signal transduction pathways in bacteria (36). *pipB* encoded translocated effector of T3SS in SPI-2 (11, 12). Moreover, *pipB* null mutant caused reduced virulence activities in bovine enteric infections (9) and related to colonization of the cecum in chickens (35). Domain Chaperone_III in *pipC* contained non-synonymous substitutions, which was involved in the type III secretion system (T3SS) and in delivering virulence effector proteins from *Salmonella* to host cells (37). The genes under positive selection could be possible targets for mutational studies (35). Due to the important role of SPI-6 in virulence of pathogenic *Salmonella* (18), three Asian strains in IIA may have different virulence activities. *S.* Gallinarum 287/91, *S.* Virchow SL491, and *S.* Paratyphi B SPB7 did not contain SPI-6 either (17). Thus, the

gain or loss of SPI-6 has occurred independently in different serovars. I could not decide the acquisition of SPI6-GI1 was introduced independently or it replaced SPI-6 between tRNA-*asp* and *sinR*. Since SPI-6 was located next to tRNA-asp and contained Rhs family proteins, both of which are associated with rearrangement or acquisition of new genetic elements (19, 38) , this location is likely to be a hot spot for recombination events and HGT. SPI-6 and *saf* phylogenetic trees showed different geographic structures from the whole genome tree, indicating geographic location played an important role in the evolution of SPI-6. According to the distribution of T6SS, *saf*, and *tcf* genes, the transfer of these gene clusters were independent events.

In addition, the *tcf* fimbrial operon, *tinR*, and *tioA* were only identified downstream of *sinR* in two strains in IIA, pepper_Vietnam and squid_Vietnam. These genes clustered together with putative transposase remnants at the downstream of SPI-6 in *S*. Typhi, but not in *S*. Typhimurium (8). Porwollik (39) demonstrated that the *Salmonella* with broad host range always had higher numbers of fimbrial operons than those with host restriction, except for *S*. Paratyphi B with a short evolutionary history. Moreover, diversification of fimbrial operon in *Salmonella* may contribute to virulence activities (34, 40). Thus, the presence of *tcf* genes may facilitate the survival of these two strains.

References

1. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.
2. **CDC.** 2006. *Salmonella* Surveillance: Annual Summary.
3. **Greene SK, Daly ER, Talbot EA, Demma LJ, Holzbauer S, Patel NJ, Hill TA, Walderhaug MO, Hoekstra RM, Lynch MF, Painter JA.** 2008. Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. Epidemiol Infect **136:**157-165.
4. **Schneider JL, White PL, Weiss J, Norton D, Lidgard J, Gould LH, Yee B, Vugia DJ, Mohle-Boetani J.** 2011. Multistate outbreak of multidrug-resistant *Salmonella* newport infections associated with ground beef, October to December 2007. J Food Prot **74:**1315-1319.
5. **Bell RL, Cao G, Meng J, Allard MW, Keys CE, Hill T, Ottesen AR, Brown EW.** 2012. *Salmonella* Newport Contamination of Produce: Ecological, Genetic, and Epidemiological Aspects, p. In press, *Salmonella*: Classification, Genetics and Disease. Nova Publishers, Hauppauge NY.
6. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010. Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

7. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

8. **Sabbagh SC, Forest CG, Lepage C, Leclerc JM, Daigle F.** 2010. So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. FEMS Microbiol Lett **305:**1-13.

9. **Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE.** 1998. Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. Mol Microbiol **29:**883-891.

10. **Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L.** 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. Mol Microbiol **39:**260-271.

11. **Knodler LA, Celli J, Hardt WD, Vallance BA, Yip C, Finlay BB.** 2002. *Salmonella* effectors within a single pathogenicity island are differentially expressed and translocated by separate type III secretion systems. Mol Microbiol **43:**1089-1103.

12. **Hensel M.** 2004. Evolution of pathogenicity islands of *Salmonella enterica*. Int J Med Microbiol **294:**95-102.

13. **Rychlik I, Karasova D, Sebkova A, Volf J, Sisak F, Havlickova H, Kummer V, Imre A, Szmolka A, Nagy B.** 2009. Virulence potential of five major pathogenicity islands (SPI-1 to SPI-5) of *Salmonella enterica* serovar Enteritidis for chickens. BMC Microbiol **9:**268.

14. **Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Lofdahl S.** 1999. Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. Mol Microbiol **33:**612-622.

15. **Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella* enterica serovar Typhi CT18. Nature **413:**848-852.

16. **Schwarz S, Hood RD, Mougous JD.** 2010. What is type VI secretion doing in all those bugs? Trends Microbiol **18:**531-537.

17. **Blondel CJ, Jimenez JC, Contreras I, Santiviago CA.** 2009. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. BMC Genomics **10:**354.

18. **Jani AJ, Cotter PA.** 2010. Type VI secretion: not just for pathogenesis anymore. Cell Host Microbe **8:**2-6.

19. **Pukatzki S, McAuley SB, Miyata ST.** 2009. The type VI secretion system: translocation of effectors and effector-domains. Curr Opin Microbiol **12:**11-17.

20. **Hood RD, Singh P, Hsu F, Guvener T, Carl MA, Trinidad RR, Silverman JM, Ohlson BB, Hicks KG, Plemel RL, Li M, Schwarz S, Wang WY, Merz**

AJ, Goodlett DR, Mougous JD. 2010. A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. Cell Host Microbe **7:**25-37.

21. **Folkesson A, Lofdahl S, Normark S.** 2002. The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. Res Microbiol **153:**537-545.

22. **Humphries AD, Raffatellu M, Winter S, Weening EH, Kingsley RA, Droleskey R, Zhang S, Figueiredo J, Khare S, Nunes J, Adams LG, Tsolis RM, Baumler AJ.** 2003. The use of flow cytometry to detect expression of subunits encoded by 11 Salmonella enterica serotype Typhimurium fimbrial operons. Mol Microbiol **48:**1357-1376.

23. **Cao G, Zhao S, Strain E, Luo Y, Timme R, Wang C, Brown E, Meng J, Allard M.** 2012. Draft genome sequences of eight *Salmonella enterica* serotype Newport strains from diverse hosts and locations. J Bacteriol **194:**5146.

24. **Lienau EK, Blazar JM, Wang C, Brown EW, Stones R, Musser S, Allard MW.** 2013. Phylogenomic Analysis Identifies Gene Gains That Define *Salmonella enterica* Subspecies I. PLoS One **8:**e76821.

25. **Timme RE, Pettengill JB, Allard MW, Strain E, Barrangou R, Wehnes C, Van Kessel JS, Karns JS, Musser SM, Brown EW.** 2013. Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. Genome Biol Evol **5:**2109-2123.

26. **Goloboff P, Farris S, Nixon K.** 2008. TNT, a program for phylogenetic analysis. Cladistics **24:**774-786.

27. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

28. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.

29. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

30. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-2739.

31. **Scheurwater E, Reid CW, Clarke AJ.** 2008. Lytic transglycosylases: bacterial space-making autolysins. Int J Biochem Cell Biol **40:**586-591.

32. **Boyd EF, Davis BM, Hochhut B.** 2001. Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. Trends Microbiol **9:**137-144.

33. **Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW.** 2012. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **13:**32.

34. **Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW.** 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One **8:**e55254.

35. **Soyer Y, Orsi RH, Rodriguez-Rivera LD, Sun Q, Wiedmann M.** 2009. Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes. BMC Evol Biol **9:**264.

36. **Zhulin IB, Nikolskaya AN, Galperin MY.** 2003. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. J Bacteriol **185:**285-294.

37. **Luo Y, Bertero MG, Frey EA, Pfuetzner RA, Wenk MR, Creagh L, Marcus SL, Lim D, Sicheri F, Kay C, Haynes C, Finlay BB, Strynadka NC.** 2001. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. Nat Struct Biol **8:**1031-1036.

38. **Hill CW.** 1999. Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. Res Microbiol **150:**665-674.

39. **Porwollik S (ed.).** 2011. *Salmonella*: from Genome to Function. Caister.

40. **Yue M, Rankin SC, Blanchet RT, Nulton JD, Edwards RA, Schifferli DM.** 2012. Diversification of the *Salmonella* fimbriae: a model of macro- and microevolution. PLoS One **7:**e38596.

TABLE III-1. General Characteristics of genes/open reading frames (ORFs) in SPI5-GI1 in *S*. Newport SL254.

ORF	Size (bp)	GC%	Best BLASTP Hit				Super Family
			Description	Source	E Value	Locus Tag	
A1129	1020	48.1	site-specific recombinase, phage integrase family	*S*. 4,[5],12:i:- CVM23701	0	EDZ14076.1	DNA_BRE_C
A1130	225	48	conserved hypothetical protein	*S*. 4,[5],12:i:- CVM23701	4e-49	EDZ14065.1	
A1131	183	48.6	conserved hypothetical protein	*S*. 4,[5],12:i:- CVM23701	6e-37	EDZ14062.1	
A1132	555	47	conserved hypothetical protein	*S*. 4,[5],12:i:- CVM23701	1e-129	EDZ14071.1	Sipho_GP157
A1133	2265	47.4	conserved hypothetical protein (Gifsy-1 prophage RecE; exodeoxyribonuclease VIII)	*S*. 4,[5],12:i:- CVM23701	0	EDZ14054.1	
A1134	246	44.7	conserved hypothetical protein	*S*. 4,[5],12:i:- CVM23701	9e-53	EDZ14075.1	
A1135	324	46.6	hypothetical protein	*S*. 4,[5],12:i:- CVM23701	5e-73	EDZ14073.1	
A1136	216	45.8	hypothetical protein	*S*. 4,[5],12:i:- CVM23701	2e-44	EDZ14068.1	
A1137	348	53.2	hypothetical protein	*S*. 4,[5],12:i:- CVM23701	2e-78	EDZ14057.1	DUF1019
A1138	225	50.7	conserved hypothetical protein	*S*. 4,[5],12:i:- CVM23701	4e-49	EDZ14066.1	
A1139	981	49	Replication protein	*S*. 4,[5],12:i:- CVM23701	0	EDZ14055.1	
A1140	1389	51.9	bacteriophage Nil2 protein P DnaB family protein	*Klebsiella pneumoniae* subsp. rhinoscleromatis ATCC 13884	0	EEW40180.1	DnaB, P-loop NTPase
A1141	681	47.6	conserved hypothetical protein	*Klebsiella pneumoniae* subsp.	1e-54	EEW40179.1	

				rhinoscleromatis ATCC 13884			
A1142	249	56.2	conserved hypothetical protein	*S.* 4,[5],12:i:- CVM23701	2e-56	EDZ15124.1	
A1143	603	49.6	adenine methylase	*S.* 4,[5],12:i:- CVM23701	1e-149	EDZ15094.1	MT-A70
A1144	339	54	conserved hypothetical protein	*S.* 4,[5],12:i:- CVM23701	3e-77	EDZ15086.1	
A1145	192	40.6	hypothetical phage-related protein	*S.* 4,[5],12:i:- CVM23701	9e-40	EDZ15059.1	
A1146	597	47.6	gifsy-2 prophage protein	*S.* 4,[5],12:i:- CVM23701	4e-145	EDZ15117.1	DUF1367
A1147	579	45.9	conserved hypothetical protein	*S.* 4,[5],12:i:- CVM23701	6e-106	EDZ15075.1	DUF1133
A1148	189	48.1	hypothetical protein	*S.* 4,[5],12:i:- CVM23701	1e-39	EDZ15091.1	
A1149	309	48.2	phage-holin analog protein	*S.* Heidelberg str. SL486	2e-70	EDZ23960.1	
A1150	540	55	lysozyme	*S.* Weltevreden str. HI_N05-537	1e-128	EDZ29335.1	lysozyme
A1151	138	42	putative bacteriophage protein	*S.* Newport SL317	6e-23	EDX48909.1	
A1152	432	56.3	bacteriophage lysis protein	*S.* arizonae serovar 62:z4,z23:-- RSK2980	2e-95	ABX22264.1	
A1153	630	51.1	conserved hypothetical protein(phage related)	*S.* 4,[5],12:i:- CVM23701	6e-148	EDZ15067.1	
A1154	1623	55.3	gp33 TerL	*S.* Paratyphi C RKS4594	0	ACN45201.1	
A1155	1437	56.7	phage-associated protein	*S.* Typhi str. E00-7866	0	ZP_03346749.1	COG3567
A1156	618	57	HI1407 hypothetical protein-like protein	*Shigella flexneri*	6e-134	AAQ07463.1	Phage_Mu_F
A1157	1233	53.6	hypothetical protein (phage related)	*S.* Paratyphi C RKS4594	0	ACN45204.1	DUF2213
A1158	498	55.8	hypothetical protein (phage related)	*S.* Paratyphi B SPB7	3e-114	ABX67616.1	

A1159	942	53	bacteriophage protein	*S*. Paratyphi B SPB7	0	ABX67615.1	Linocin_M18
A1160	390	51.8	bacteriophage protein	*S*. Typhi Ty2	2e-72	AAO69505.1	
A1161	408	53.4	bacteriophage protein	*S*. Typhi Ty2	7e-94	AAO69504.1	
A1162	555	52.4	bacteriophage protein	*S*. Typhi Ty2	1e-129	AAO69503.1	
A1163	390	55.1	bacteriophage protein	*S*. Typhi Ty2	4e-87	AAO69502.1	
A1164	564	56.2	bacteriophage protein	*S*. Typhi Ty2	2e-105	AAO69501.1	
A1165	1146	54.3	bacteriophage protein	*S*. Paratyphi C RKS4594	0	ACN45212.1	DUF3383
A1166	444	55.2	bacteriophage protein	*S*. Paratyphi C RKS4594	1e-102	ACN45213.1	DUF3277
A1167	453	54.3	putative bacteriophage protein	*S*. 4,[5],12:i:- CVM23701	2e-102	EDZ15105.1	
A1168	2010	56.5	lytic transglycosylase, catalytic	*Enterobacter* sp. 638	0	ABP59494.1	lysozyme
A1169	576	54.9	putative bacteriophage protein	*Enterobacter* sp. 638	2e-122	ABP59495.1	
A1170	303	54.1	bacteriophage protein	*S*. Typhi CT18	1e-62	NP_455541.1	
A1171	1068	55.4	putative bacteriophage protein	*Enterobacter* sp. 638	0	ABP59497.1	
A1172	165	47.9	putative secreted protein	*E coli* F11	4e-30	EDV66657.1	
A1173	456	41.9	hypothetical protein	*Erwinia* sp. Ejp617	3e-83	ADP12360.1	
A1174	753	55.2	hypothetical protein(phage related)	*E coli* IAI1	5e-169	CAQ99465.1	Phage_base_V
A1175	354	51.1	putative bacteriophage protein	*S*. Typhi E01-6750	3e-72	ZP_03354527.1	
A1176	1200	52.5	bacteriophage protein	*S*. Typhi Ty2	0	AAO69487.1	Baseplate_J
A1177	681	48.3	putative bacteriophage protein	*S*. 4,[5],12:i:- CVM23701	4e-160	EDZ15103.1	DUF2612
A1178	1533	50.7	bacteriophage tail protein	*S*. Typhi Ty2	0	AAO69485.1	
A1179	528	43.2	Phage tail assembly chaperone gp38	*S*. Choleraesuis A50	4e-114	EFZ05846.1	Caudo_TAP
A1180	126	43.7	hypothetical protein	*S*. Typhi E01-6750	2e-13	ZP_03350039.1	
A1181	639	47.9	hypothetical protein (phage tail)	*S*. Paratyphi A AKU_12601	1e-83	CAR59077.1	Caudo_TAP

A1182	393	51.1	hypothetical protein	*S.* Paratyphi B SPB7	5e-68	ABX66451.1	DUF1353
A1183	420	30.7	hypothetical protein	*S.* 4,[5],12:i:- CVM23701	1e-96	EDZ15089.1	

TABLE III-2. General characteristics of genes/open reading frames (ORFs) of SPI5-GI2 in strain shrimp_India.

ORF	Size (bp)	GC%	Best BLASTP Hit				Super Family
			Description	Source	E Value	Locus Tag	
443_05447	1023	45.8	integrase protein	*Escherichia albertii* TW07627	0	EDS90736.1	DNA_BRE_C (phage integrase)
443_05442	237	47.3	conserved hypothetical protein (phage integrase)	*E. coli* O157:H7 EC869	1e-28	ZP_02812412.1	
443_05437	429	56.4	hypothetical protein	*S.* Gallinarum 287/91	4e-100	CAR37063.1	
443_05432	2145	55.3	C-5 cytosine-specific DNA methylase	*S.* Saintpaul SARA29	0	EDZ10721.1	AdoMet_MTa-ses
443_05427	258	52.7	conserved hypothetical protein	*S.* Saintpaul SARA29	9e-54	EDZ10747.1	
443_05422	180	41.1	conserved hypothetical protein	*S.* Kentucky CDC 191	7e-36	EDZ19030.1	DUF1187
443_05417	843	53	hypothetical protein (DNA adenine methyltransferase)	*Providencia alcalifaciens* DSM 30120	5e-130	EEB47035.1	MethyltransfD12
443_05412	234	48.7	hypothetical protein	*S.* Choleraesuis SC-B67	3e-67	AAX64235.1	
frameshift	830	54.3	protein RecT	*S.* Weltevreden HI_N05-537	4e-149	EDZ29450.1	RecT
443_05397	247	51	exodeoxyribonuclease 8	*S.* Saintpaul SARA29	3e-50	EDZ10745.1	
443_12533	2118	50	exodeoxyribonuclease 8	*S.* Newport SL317	0	EDX48971.1	Exonuc_VIII
443_12538	336	42.9	LygB	*S.* Newport SL317	7e-77	EDX49023.1	
443_12543	207	40.6	prophage Kil protein	*S.* Saintpaul SARA29	9e-44	EDZ09313.1	kil
443_12548	276	45.7	hypothetical protein	*S.* Gallinarum 287/91	3e-57	CAR37068.1	

443_12553	309	49.8					
443_12558	306	45.4	hypothetical protein	*E. coli* TA280	2e-06	EGI41136.1	
443_12563	402	43.5	HTH-type transcriptional regulator DicA	*S.* Saintpaul SARA29	8e-91	EDZ09310.1	HTH_XRE
443_12568	219	47.5	DNA-binding transcriptional regulator DicC	*E. coli* B185	3e-31	EFF07713.1	PRK09744
443_12573	495	55.6	conserved hypothetical protein	*S.* Newport SL317	1e-115	EDX48223.1	DUF1019
443_12578	903	52.9	conserved hypothetical protein	*S.* Kentucky CDC 191	2e-76	EDZ19051.1	DUF1376
443_12583	753	55.9	DNA replication protein dnaC	*E. coli* UMNF18	1e-140	AEJ55928.1	P-loop NTPase
443_12588	396	51.8	phage encoded DNA-binding protein	*S.* Gallinarum 287/91	1e-92	CAR37075.1	DUF977
443_12593	273	48.4	conserved hypothetical protein	*S.* Kentucky CDC 191	2e-59	EDZ19038.1	
443_12598	156	53.2	prophage maintenance protein	*E. coli* O127:H6 E2348/69	2e-25	CAS08950.1	HOK_GEF
443_12603	195	43.6	conserved hypothetical protein	*S.* Saintpaul SARA29	8e-39	EDZ09306.1	
443_12608	438	51.6	NinB protein	*S.* Saintpaul SARA29	4e-103	EDZ09302.1	NinB
443_12613	282	52.5	conserved hypothetical protein	*S.* Weltevreden HI_N05-537	1e-63	EDZ29331.1	DUF1364
443_12618	537	46.4	putative bacteriophage protein	*S.* Typhimurium 14028S	6e-111	ACY87930.1	DUF1133

443_12623	177	48.6	hypothetical protein	*Photorhabdus luminescens* subsp. *laumondii* TTO1s	1e-29	CAE15865.1	YcfA
443_12628	408	49.5	conserved hypothetical protein	*S.* Saintpaul SARA29	1e-94	EDZ09559.1	UPF0150,HTH_XRE
443_12633	303	47.9	phage-holin analog protein	*S.* Newport SL254	2e-69	ACF64184.1	
443_12638	540	54.6	lysozyme	*S.* Weltevreden HI_N05-537	4e-129	EDZ29335.1	Lysozyme_like
443_12643	99	44.4	putative bacteriophage protein	*S.* Heidelberg SL476	2e-14	ACF68737.1	
443_12648	465	55.9	bacteriophage lysis protein	*S.* Weltevreden HI_N05-537	3e-95	EDZ29378.1	Phage_lysis
443_12653	183	47	conserved hypothetical protein	*S.* Saintpaul SARA29	2e-32	EDZ10296.1	DUF826
443_12658	630	51.3	conserved hypothetical protein	*S.* Weltevreden HI_N05-537	4e-147	EDZ29428.1	
443_12663	1620	50.7	gp33 TerL	*S.* Saintpaul SARA29	0	EDZ09552.1	
443_12668	1521	53.5	phage-associated protein, HI1409 family	*S.* Weltevreden HI_N05-537	0	EDZ29453.1	COG3567
443_12673	690	56.1	gp16	*S.* Weltevreden HI_N05-537	1e-167	EDZ29333.1	Phage_Mu_F
443_12678	1347	56.2	conserved hypothetical protein	*S.* Saintpaul SARA29	0	EDZ09564.1	DUF2213
443_12683	483	59.2	conserved hypothetical protein (phage related)	*S.* Saintpaul SARA29	1e-110	EDZ09554.1	
443_12688	1029	54.5	gp12	*S.* Weltevreden HI_N05-537	0	EDZ29350.1	
443_12693	348	56.3	conserved hypothetical	*S.* Saintpaul SARA29	3e-79	EDZ09565.1	

			protein				
443_12698	456	59	hypothetical protein	*S.* Saintpaul SARA29	7e-108	EDZ09560.1	
443_12703	585	49.2	hypothetical protein	*S.* Saintpaul SARA29	5e-139	EDZ09547.1	
443_12708	366	55.2	conserved hypothetical protein	*S.* Saintpaul SARA29	5e-85	EDZ09551.1	
443_12713	546	53.8	conserved hypothetical protein	*S.* Saintpaul SARA29	6e-131	EDZ09405.1	
443_12718	1485	53.4	conserved hypothetical protein	*S.* Saintpaul SARA29	0	EDZ09406.1	
443_12723	447	51.5	conserved hypothetical protein	*S.* Saintpaul SARA29	3e-106	EDX48993.1	
443_12728	405	52.3	conserved hypothetical protein	*S.* Newport SL317	3e-91	EDX48993.1	
443_12733	183	51.4	gp15' E'	*S.* Weltevreden HI_N05-537	1e-36	EDZ29398.1	
443_12738	2172	55	transglycosylase SLT domain protein	*S.* Newport SL317	0	EDX49051.1	Lysozyme_like
443_12743	711	51.8	conserved hypothetical protein	*S.* Saintpaul SARA29	5e-170	EDZ09522.1	
443_12748	303	53.5	conserved hypothetical protein	*S.* Saintpaul SARA29	1e-68	EDZ09526.1	
443_12753	870	53.7	conserved hypothetical protein	*S.* Newport SL317	0	EDX49041.1	
443_12758	678	53.7	phage P2 baseplate assembly protein gpV	*S.* Weltevreden HI_N05-537	5e-163	EDZ29362.1	Phage_base_V
443_12763	357	53.5	conserved hypothetical	*S.* Newport SL317	4e-81	EDX48972.1	

			protein				
443_12768	1242	52.7	conserved hypothetical protein	*S.* Weltevreden HI_N05-537	0	EDZ29403.1	Baseplate_J
443_12773	603	50.4	hypothetical bacteriophage protein	*S.* Newport SL317	8e-145	EDX48931.1	DUF2612
443_12778	1452	53.4	side tail fiber protein	*S.* Newport SL317	0	EDX49031.1	Collar
443_12783	825	48.7	UPF0189 protein LA_4133	*S.* Weltevreden 2007-60-3289-1	0	CBY95370.1	Macro
443_12788	570	48.9	phage tail assembly protein	*S.* Saintpaul SARA29	1e-123	EDZ13156.1	Caudo_TAP
443_12793	855	39.8	conserved hypothetical protein	*S.* Saintpaul SARA29	0	EDZ09975.1	HtrL_YibB
443_12798	1541	35.7	putative sulfatase domain protein	*S.* Saintpaul SARA29	0	EDZ09946.1	Sulfatase

TABLE III-3. Average pairwise distance (no. of nucleotide difference) of SPI-5 and *saf* fimbrial operon for *S*. Newport and outgroups genomes.

Average pairwise distance of SPI-5							
	Tennessee	Kentucky	Dublin	Newport III	Gallinarum	Typhimurium	
Tennessee							
Kentucky	62(6)						
Dublin	76(6)	73(6)					
Newport III	70(6)	71(6)	18(4)				
Gallinarum	88(6)	83(6)	50(6)	50(6)			
Typhimurium	83(6)	82(6)	41(5)	31(5)	41(5)		
Newport II	81(6)	74(6)	49(6)	40(5)	35(5)	18(4)	
Average pairwise distance of *saf* operon							
	Newport IIA&B	Shrimp_India	Tennessee	Pig_ear_CA	Newport III	Typhimurium	Newport IIC
Newport IIA&B							
Shrimp_India	79 (7)						
Tennessee	71 (7)	21 (4)					
Pig_ear_CA	210 (9)	212 (9)	217 (9)				
Newport III	210 (9)	210 (9)	215 (9)	11 (3)			
Typhimurium	137 (9)	141 (9)	144 (9)	197 (9)	197 (9)		
Newport IIC	223 (9)	221 (9)	222 (9)	108 (8)	109 (8)	194 (9)	
Gallinarum	233 (8)	236 (6)	241 (9)	105 (8)	104 (8)	193 (9)	67 (7)

Distances were calculated using the concatenated alignment of SNPs of SPI-5 and *saf* operon that estimate the diversity between two major lineages and outgroup genomes observed. Standard deviation was listed in parentheses.

TABLE III-4. Single Nucleotide Polymorphisms (SNPs) of SPI-5 genes defining *S*. Newport lineages II and III.

Gene	*S*. Newport SL254 (lineage II)	*S*. Newport SL317 (lineage III)	Nuc	AA	Position
pipA	SNSL254_A1184	SNSL317_A1439	G->A	D/N	70
pipA	SNSL254_A1184	SNSL317_A1439	A->C	R/P	328
pipA	SNSL254_A1184	SNSL317_A1439	G->C	R/P	329
pipA	SNSL254_A1184	SNSL317_A1439	T->C	V/A	485
pipB	SNSL254_A1185	SNSL317_A1440	A->G	N/D	217
pipB	SNSL254_A1185	SNSL317_A1440	A->C	D/A	308
pipB	SNSL254_A1185	SNSL317_A1440	T->C	S	369
pipB	SNSL254_A1185	SNSL317_A1440	A->C	K/Q	412
pipB	SNSL254_A1185	SNSL317_A1440	A->G	K/D	517
pipB	SNSL254_A1185	SNSL317_A1440	A->C	K/D	519
pipB	SNSL254_A1185	SNSL317_A1440	A->G	N/D	532
pipB	SNSL254_A1185	SNSL317_A1440	C->A	T/N	554
pipB	SNSL254_A1185	SNSL317_A1440	T->A	L	558
sopB	SNSL254_A1187	SNSL317_A1442	A->G	A	114
sopB	SNSL254_A1187	SNSL317_A1442	T->C	S/P	127
sopB	SNSL254_A1187	SNSL317_A1442	T->C	V/A	134
sopB	SNSL254_A1187	SNSL317_A1442	T->C	D	480
sopB	SNSL254_A1187	SNSL317_A1442	T->C	G	1473
pipC	SNSL254_A1186	SNSL317_A1441	A->C	D/A	29

pipC	SNSL254_A1186	SNSL317_A1441	G->A	A	48
pipC	SNSL254_A1186	SNSL317_A1441	G->T	L	66
pipC	SNSL254_A1186	SNSL317_A1441	T->C	L	69
pipC	SNSL254_A1186	SNSL317_A1441	G->T	L	186
pipC	SNSL254_A1186	SNSL317_A1441	C->T	Y	192
pipC	SNSL254_A1186	SNSL317_A1441	A->G	T/A	193
pipD	SNSL254_A1189	SNSL317_A1444	C->T	D	189
pipD	SNSL254_A1189	SNSL317_A1444	A->G	E	216
pipD	SNSL254_A1189	SNSL317_A1444	A->G	I/V	259
pipD	SNSL254_A1189	SNSL317_A1444	G->A	A	345
pipD	SNSL254_A1189	SNSL317_A1444	T->C	Y	435
pipD	SNSL254_A1189	SNSL317_A1444	C->T	A/V	509
pipD	SNSL254_A1189	SNSL317_A1444	C->T	F	822
pipD	SNSL254_A1189	SNSL317_A1444	C->T	A/V	920
pipD	SNSL254_A1189	SNSL317_A1444	A->G	T	966
pipD	SNSL254_A1189	SNSL317_A1444	A->G	K	969
pipD	SNSL254_A1189	SNSL317_A1444	T->C	I	972
pipD	SNSL254_A1189	SNSL317_A1444	A->C	R	1002
pipD	SNSL254_A1189	SNSL317_A1444	C->T	P	1068
pipD	SNSL254_A1189	SNSL317_A1444	G->A	S	1095

A total of 39 SNPs in five genes in SNP-5 were identified. They defined *S*. Newport lineages II and III. All of these informative SNPs could be used as potential biomarkers to differentiate strains during outbreak trace-back investigations. There are total 18 SNPs causing non-synonymous substitutions.

TABLE III-5. Characteristics of genes/open reading frames (ORFs) in SPI6-GI1 in *S.* Virchow SL491.

ORF	Size (bp)	GC%	Best BLASTP hit				Super family
			Description	Source	E Value	Locus Tag	
SeV_A1941	1539	35.8	hypothetical protein	*S.* Paratyphi B SPB7	0	ABX68679.1	NA
SeV_A1942	123	43.9	hypothetical protein	*S.* Virchow SL491	6e-22	EDZ00328.1	NA
SeV_A1943	876	59.5	putative small GTP-binding domain protein	S. Paratyphi B SPB7	0	ABX68684.1	Ras_like_GTPase
SeV_A1944	189	50.3	hypothetical protein	*S.* Virchow SL491	1e-37	EDZ00604.1	NA
SeV_A1945	714	55.2	transcriptional regulator	*S.* Paratyphi B SPB7	2e-173	ABX68686.1	NA
SeV_A1946	567	55.4	conserved hypothetical protein	*S.* Paratyphi B SPB7	5e-140	ABX68687.1	NA
SeV_A1947	453	52.8	conserved hypothetical protein	*S.* Gallinarum 287/91	1e-109	CAR36187.1	NA
SeV_A1948	453	51.4	hypothetical protein	*S.* Gallinarum 287/91	2e-109	CAR36186.1	NA
SeV_A1949	540	56.3	conserved hypothetical protein	*S.* Gallinarum 287/91	4e-126	CAR36185.1	NA
SeV_A1950	414	43.5	conserved hypothetical protein	*S.* Gallinarum 287/91	1e-97	CAR36184.1	NA
SeV_A1951	276	49.6	YkfF protein	*S.* Gallinarum 287/91	6e-64	CAR36183.1	DUF905
SeV_A1952	819	56.3	protein YafZ	*S.* Gallinarum 287/91	0	CAR36182.1	DUF932
SeV_A1953	222	60.4	hypothetical protein	*S.* Virchow SL491	3e-45	EDZ02031.1	NA
SeV_A1954	384	57.3	hypothetical protein	*S.* Gallinarum 287/91	1e-91	CAR36180.1	NA
SeV_A1955	459	61.2	hypothetical protein	*S.* Paratyphi B SPB7	3e-110	ABX68694.1	Antirestrict
SeV_A1956	480	57.9	DNA repair protein RadC	*S.* Gallinarum 287/91	3e-111	CAR36178.1	MPN
SeV_A1957	291	53.3	putative cytoplasmic protein	*S.* Gallinarum 287/91	4e-65	CAR36177.1	NA

SeV_A1958	222	59.5	conserved domain protein	*S.* Gallinarum 287/91	1e-48	CAR36176.1	DUF987
SeV_A1959	369	57.7	YagBYeeUYfjZ family protein	*S.* Gallinarum 287/91	2e-86	CAR36175.1	YagB_YeeU_YfjZ
SeV_A1960	372	58.3	Aec75	*S.* Gallinarum 287/91	4e-88	CAR36174.1	YagB_YeeU_YfjZ
SeV_A1961	645	56.4	conserved hypothetical protein	*S.* Gallinarum 287/91	8e-159	CAR36173.1	NA
SeV_A1962	378	57.7	conserved hypothetical protein	*S.* Paratyphi B SPB7	1e-88	ABX68701.1	DUF1219
SeV_A1963	423	54.8	YeeW protein	*S.* Paratyphi B SPB7	4e-98	ABX68702.1	NA
SeV_A1964	195	51.3	Aec78	*S.* Paratyphi B SPB7	7e-40	ABX68703.1	DUF957
SeV_A1965	846	50.2	Aec79	*S.* Paratyphi B SPB7	0	ABX68704.1	NA

FIG III-1. Whole genome parsimony tree of *S*. Newport and 11 outgroup genomes.

S. Newport strains showed identical phylogenies as a previous study. There are six equally most parsimonious trees identified with a length of 209114, consistency index (CI) of 0.616, and retention index (RI) of 0.888. Two gene clusters encoding bacteriophage genes were identified, named as SPI5-GI1 and SPI5-GI2.

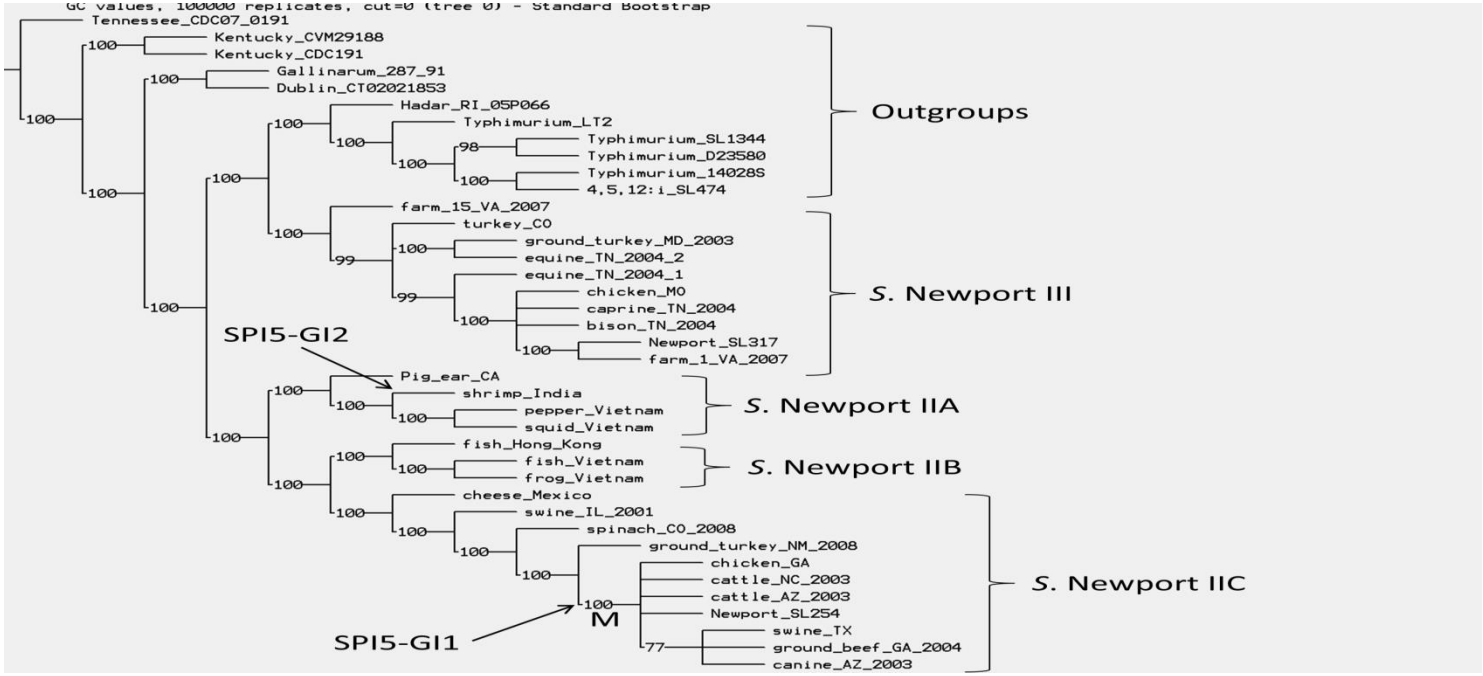


FIG III-2. Genetic organizations of SPI5-GI1 and SPI5-GI2.

S. Newport SL254 and strain shrimp_India were selected to display contents of SPI5-GI1 and SPI5-GI2. Both SPI5-GIs were inserted between tRNA-*Ser* and *pipA*.

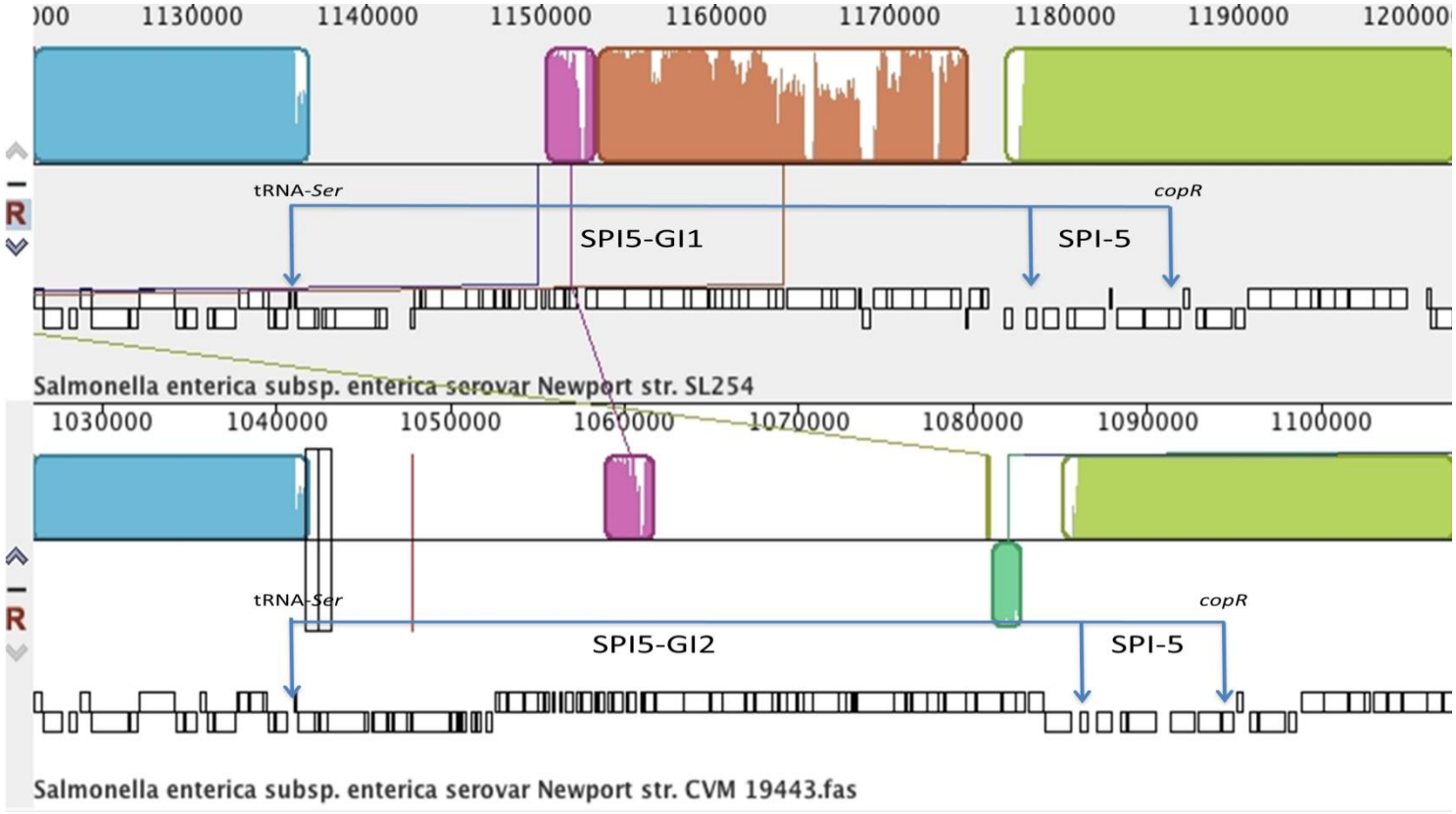


FIG III-3. Parsimony tree of SPI-5 genes.

There are 227 equally most parsimonious trees identified with a length of 187, and consistency index (CI) of 0.797, and retention index (RI) of 0.942. Lineages II and III were separated by outgroup genomes. Lineage II displayed close relationship with *S.* Typhimurium group, *S.* 4,[5],12:i:- SL474, *S.* Hadar RI_05P066, and *S.* Gallinarum 287/91; lineage III showed close relationship with *S.* Dublin CT_02021853.

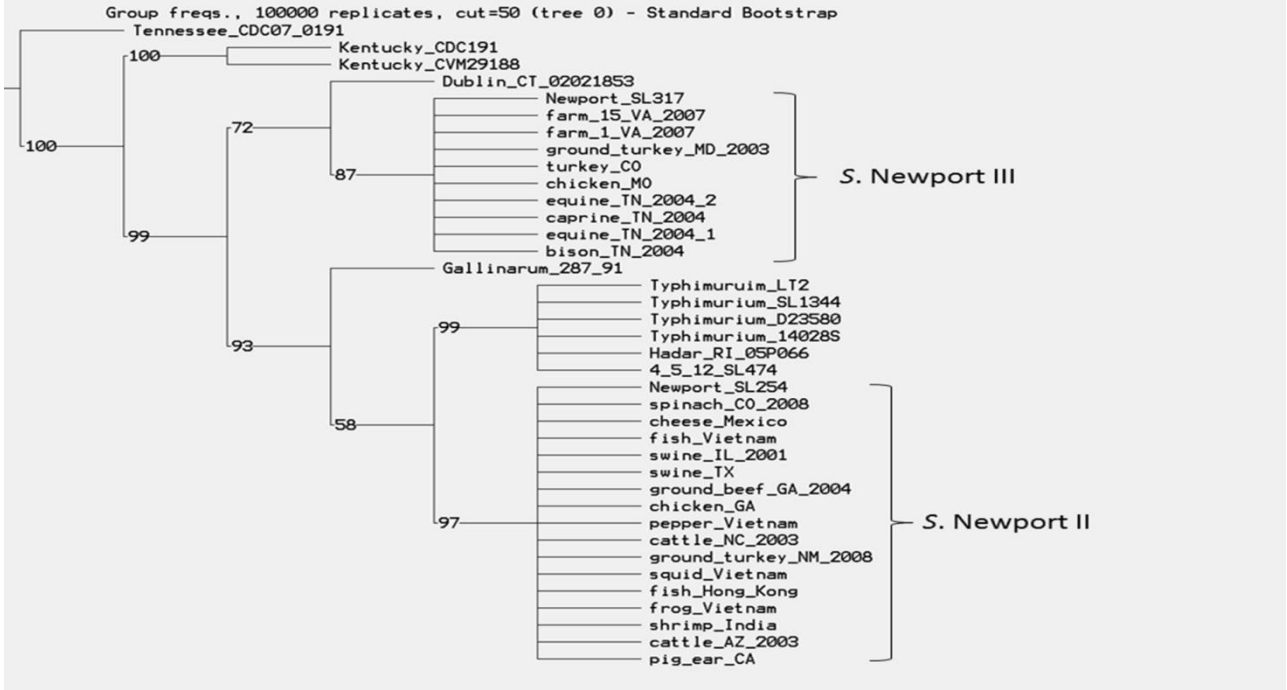


FIG III-4. Genetic organizations of SPI-6 and SPI6-GI1.

S. Newport SL317 and strain pepper_Vietnam were selected.

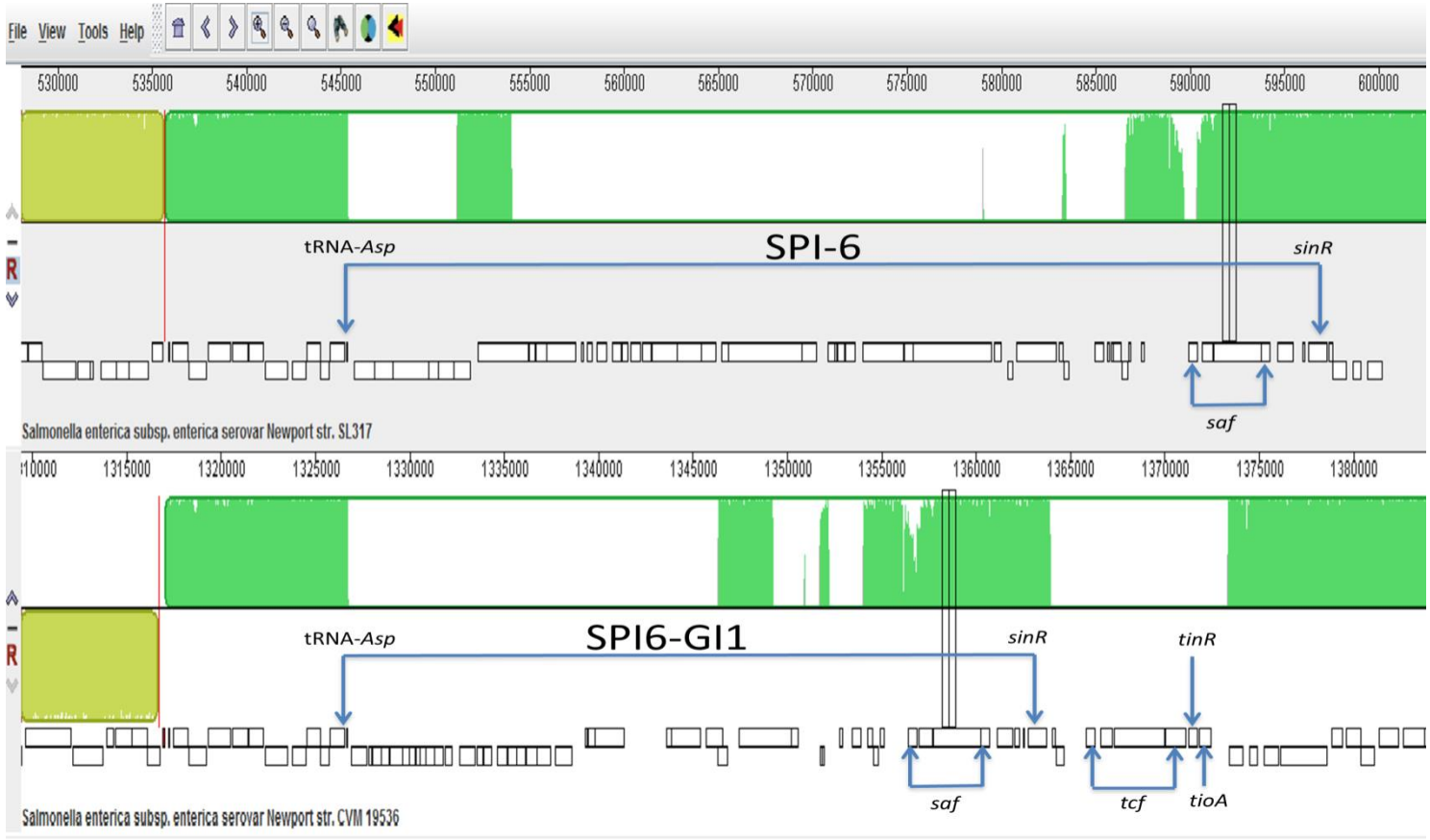


FIG III-5. Parsimony tree of SPI-6 genes.

There were 208 equally most parsimonious trees determined with a length of 1029, consistency index (CI) of 0.914, and retention index (RI) of 0.984. SPI-6 tree displayed clear geographic structure. The Asian strains were decoupled from all American strains.

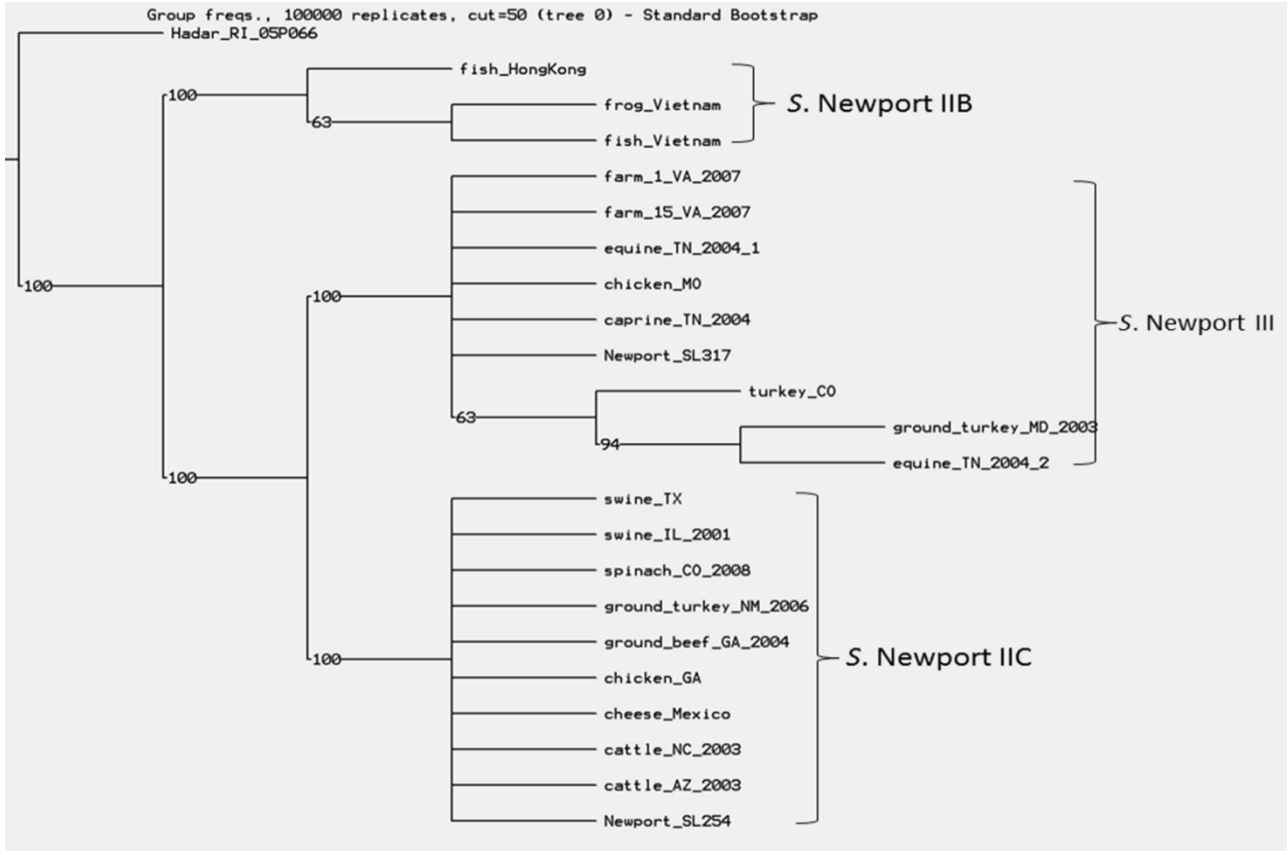
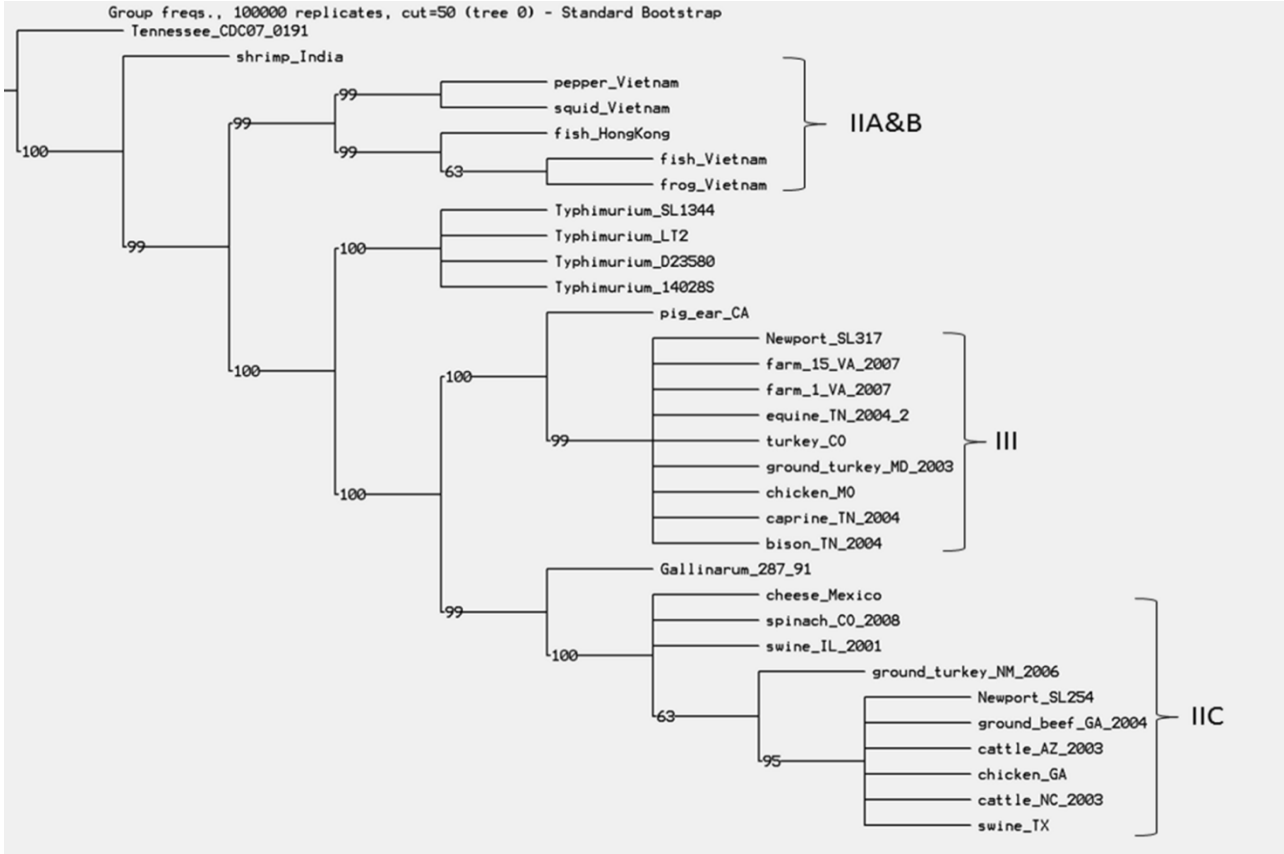


FIG III-6. Parsimony phylogenetic tree of *saf* gene cluster.

There are 210 equally most parsimonious trees determined with a length of 493, consistency index (CI) of 0.840, and retention index (RI) of 0.970. The *saf* cluster had clear geographic structure, meaning that all Asian strains were decoupled from those from the Americas.



CHAPTER IV: COMPARATIVE GENOMICS OF MULTIDRUG RESISTANT STRAINS OF *S.* NEWPORT

Abstract

Salmonella enterica subspecies *enterica* serotype Newport (*S*. Newport) infections have increased in North America since 1996. Multidrug resistant *S*. Newport has been an emerging concern of public health. The IncA/C plasmids were thought to be responsible for the dissemination of MDR *S*. Newport strains. Therefore, it is necessary to characterize the *S*. Newport MDR strains. A total of eight *S*. Newport genomes were selected to perform comparative genomics, including seven MDR strains. Complete and partial sequences of plasmids were determined in compared genomes, including genes in plasmid pSN254. Resistance genes were identified, including *floR, tetAR, strAB,* and *merABDEPRT*. The study discovered that all *S*. Newport MDR strains harbored *cas3* as pseudogene containing one premature stop codon. In contrast, susceptible strain carried intact CRISPR/cas system. The unfunctional CRISPR/cas system may facilitate the acquisition of IncA/C plasmids.

Introduction

Prevalence of antimicrobial resistance among *Salmonella* has been important issue of food safety and clinical therapy (1). Multidrug resistant *S.* Newport strains are emerging concerns for public health in North America (2, 3). Strains known as Newport MDRAmpC are resistant to ampicillin (*pse-1*), chloramphenicol (*floR*), streptomycin (aadA2), sulfamethoxazole (*sul1*), and tetracycline (*tet*) (ACSSuT) (4, 5). *S.* Newport SL254 possessed one bla_{CMY-2} positive plasmid named pSN254, contributing to antimicrobial resistance type ACSSuT (6). The genes responsible to resistance also exhibited in *Salmonella* Genomic Island 1 (SGI-1) (7).

Chapter II data indicated that seven MDR strains from different sources and locations formed node M, all of which belonged to resistance type ACSSuT with additional resistance to different antibiotics. Phylogenetic analysis reported that these seven strains showed close relationship. Therefore, it is necessary to investigate the common genetic traits of node M. Strain or node specific gene clusters were identified to illustrate different genomic contents of the compared genomes. Besides the diversity in *Salmonella* pathogenicity islands (8), different categories of genes related to virulence were examined, including non-coding RNAs and fimbrial operons. Non-coding RNAs are those non-translated nucleotides and encoded in intergenic regions, which play important roles in gene regulations and expression under different stress or virulence conditions (9). A total of 106 non-coding RNAs were selected from previous studies (10-12). Fimbriae in *Salmonella* play important roles in pathogenicity during interactions and attachment between bacteria and host cells (13). Variations of different fimbrial operons may influence the virulence of different strains (14).

Clustered regularly interspaced short palindromic repeats (CRISPR)/cas system was identified as an immune system against foreign genetic elements in 90% of archaea and 40% of bacteria genomes (15). CRISPR/cas system consists of conserved short repeats that separated by distinct foreign elements called spacers, associated proteins called *cas* genes and leader sequence (16). Moreover, CRISPR/cas system was considered to associate with MDR characteristics of bacteria (17). Therefore, it is meaningful to examine CRISPR/cas get insights into the possible association between CRISPR and MDR phenotypes in *S.* Newport.

The objectives of the current chapter is (i) to determine the common and specific genetic contents of MDR strains in node M, and (ii) to investigate the distribution of plasmids in the target strains.

Materials and Methods

Genomes

A total of eight strains from diverse sources and locations were selected (Table IV-1). They are *S.* Newport CVM N18486 (AHTY00000000), *S.* Newport CVM 21538 (AHTV00000000), *S.* Newport CVM 21550 (AHTT00000000), *S.* Newport CVM 22513 (AHTU00000000), *S.* Newport CVM 22425 (AHTW00000000), *S.* Newport CVM 22462 (AHTX00000000), *S.* Newport CVM N1543 (AHTZ00000000), and *S.* Newport SL254 (ABEN01000000).

Phylogenetic analysis

Multiple genome alignment was performed using progressiveMauve (18). Over 5,000 single nucleotide polymorphisms (SNPs) were identified to construct parsimonious phylogenetic tree using tree analysis using new technology (TNT) (19) with finding minimum length 20 times and 100,000 bootstrap replications. Pairwise distance matrix was calculated using MEGA 5.10 (20) with 1,000 bootstrap iterations.

CRISPR/*cas* identification

CRISPRFinder (21) was used to identify CRISPR spacer arrays. *cas* gene cluster of *S.* Newport SL254 was used as the reference (SNSL254_A3147 to SNSL254_A3154, 5' to 3') and locus tags of *cas* gene clusters were identified using Mauve (22).

Identification of genes associated with antimicrobial resistance and virulence

Standalone blastn (23) (Blast 2.27+) was used to construct a local database and to search homologs of query genes with 90% identities as threshold.

Results

Phylogenetic analysis

Multiple genome alignment of eight genomes identified over 5,000 SNPs. Parsimonious phylogenetic tree (Fig IV-1) was constructed to display evolutionary relatedness (tree length: 5129, consistency index: 0.994, retention index: 0.616). Susceptible strain ground_turkey_NM_2008 was selected as the outgroup genome. The branch lengths (Fig IV-1, shown in brackets) indicated that strain cannine_AZ_2003 had large amount of changes in the branch, followed by strains ground_beef_GA_2004 and swine_TX. Pairwise distance matrix showed the average base differences among eight genomes (Table IV-2). Four strains displayed close relationship compared to each other with base difference at most 38 SNPs (Table IV-2).

Although the current data could not illustrate the subtyping of plasmids or identify the resistance genes to different antibiotics, the phylogenetic tree and the antibiotic resistance profiles provided important information of the genetic flux (Fig IV-1). For example, the common ancestor of node M acquired genes resistant to nine antibiotics (AMC, AMP, FOX, TIO, AXO, CHL, STR, SUL, and TET) and four strains kept the same antimicrobial resistance profiles, including swine_TX, canine_AZ_2003, ground_beef_GA_2004, and cattle_AZ_2003. *S.* Newport SL254 gained genes resistant to CAZ and GEN and loss the genes resistant to FOX, TIO, and AXO. Similarly, cattle_NC_2003 acquired gene resistant to KAN; chicken_GA loss the genes resistant to AMC, AMP, FOX, TIO, and AXO.

CRISPR/cas system

CRISPR/cas system components were examined in all compared genomes. *cas* gene cluster composing of eight genes were identified in all strains. All *cas3* in node M were identified as pseudogenes. For example, sequence alignment between strains ground_turkey_NM_2008 and *S*. Newport SL254 indicated that one substitution (C to T at position 1648) caused a premature stop codon. In node M strains, *cas* genes remained conserved nucleotide sequence with exception. For example, frameshift were identified in *cse1*, *cse4*, and *cas3* in canine_AZ_2003.

CRISPR spacer arrays were determined and *S*. Newport SL254 was used as reference genome. Strains chicken_GA, cattle_AZ_2003, and cattle_NC_2003, and swine_TX shared common spacer arrays as *S*. Newport SL254, although some spacers were missing. Moreover, various spacers were identified in ground_beef_GA_2004 and ground_turkey_NM_2008, both of which contained common spacers in *S*. Newport SL254. However, no spacer was identified in strain canine_AZ_2003 except two questionable spacers.

Plasmid sequence identification

Complete and partial nucleotide sequences of various plasmids were identified in selected strains including those contributed to antimicrobial resistance. For example, plasmid pSN254 was identified in canine_AZ_2003 and ground_beef_GA_2004, besides *S*. Newport SL254 (Figure IV-3). A total of 166 and 131 genes in pSN254 were determined in these two draft genomes, respectively. Genes contributing to antimicrobial resistance and mercury resistance were determined, including *floR*, *tetAR*, *strAB*, and *merABDEPRT*. No gene in pSN254 was identified in other selected genomes based on draft genome data. Moreover, SGI-1 was examined because genes responsible for

antimicrobial resistance type ACSSuT also exhibited in *Salmonella* genomic island 1 (SGI-1) (24). However, no insertion was identified between genes *thdF* (SNSL254_A4127) and *yidY* (SNSL254_A4128) in all eight strains.

Moreover, additional *Salmonella* plasmids were identified in different isolates sequenced. For example, pSN254_3 was determined in cattle_AZ_2003, canine_AZ_2003, and swine_TX identified, which was first identified in *S*. Newport SL254. Plasmids from *S*. Heidelberg (pSL476_3) and *S*. Bardo (pSBardoKan) were identified in ground_beef_GA_2004 and cattle_NC_2003, respectively. No plasmid sequence was identified in draft genome data of chicken_GA and ground_turkey_NM_2008.

Identification of non-coding RNA

A total of 106 non-coding RNAs (ncRNAs) were selected to examine their presence in selected genomes to investigate the possible differences in virulence and gene regulation among different strains (Table IV-4). There were 101 ncRNAs being conserved in all selected strains. Other ncRNAs were identified with multiple copies or determined in certain strains. For example, *STnc60*, *STnc190*, and *STnc290* existed in all strains but canine_AZ_2003 with differences. One T insertion was identified in position 132 in *STnc190* in ground_turkey_ NM and there are A and T insertion identified in positions 101 and 110 in *STnc190* in ground_beef_GA, respectively. Similarly, point mutations were identified in *STnc60* in five genomes (Table IV-5). Moreover, *isrB1* and *isrB2* were identified in all genomes but with two copies in five genomes, including *S*. Newport SL254, chicken_GA, swine_TX, cattle_AZ_2003, and cattle_NC_2003. All these genomes except chicken_GA showed close relationship according to phylogenetic analysis.

Discussion

S. Newport is polyphyletic consisting of three lineages based on multilocus sequencing typing analysis (25). Previous chapters (8) showed that *S.* Newport harbored extensive genetic diversities including both indels and SNPs. *S.* Newport lineages II and III displayed divergent phylogenies and evolved largely independently since their separation. Therefore, lineage specific genetic traits may exist and could play important roles in virulence, host adaptations, or antimicrobial resistance. Moreover, lineage II showed more diverse structure compared to lineage III, consisting of three subgroups and displaying clear geographic structure (8). Interestingly, all MDR strains belonged to subgroup IIC and were isolated from North America. Thus, it is meaningful to investigate the possible common genetic characterizations shared by these clinically important strains.

In the current chapter, all MDR strains showed ACSSuT phenotypes and originated from a common ancestor. Since *Salmonella* genomic island 1 was not identified in these genomes (7), IncA/C plasmids would contribute to the multidrug resistances. Complete and partial sequences of plasmids were determined in the draft genome data. Importantly, a total of 166 and 131 genes in plasmid pSN254 were identified in genome of canine_AZ_2003 and ground_beef_GA_2004, including those responsible to antimicrobial resistance (*floR*, *tetAR*, and *strAB*) and genes resistant to mercury ions (*merABDEPRT*). Thus, it is necessary to perform further genome sequencing to obtain plasmids sequence of these MDR strains and to improve the quality of the chromosomal data.

Plasmid pSN254 shared a common backbone with plasmid in *Yersinia pestis* IP275 (pIP1202) and *E. coli* D7-3 (pRAx), which is distributed broadly in MDR zoonotic pathogens (6, 26). Because the acquisition of IncA/C plasmid in *S.* Newport was a recent event (6), the mutation in *cas3* also may have happened recently. The IncA/C plasmid pSN254 was not transferable to the *E. coli* recipient strain (6, 26). Therefore, we hypothesized that strains in node M may acquire the resistant plasmid directly from their common ancestor. There is no obvious correlation between the self-transferability and plasmid components and certain chromosomal elements or unknown plasmid factors may contribute to the limitation (26).

In addition, the phylogenetic analysis and antimicrobial resistance profile demonstrated the genetic flux in the node M. For example, four strains kept the same resistance profiles (canine_AZ_2003, ground_beef_GA_2004, swine_TX, and cattle_AZ_2003) and three genomes gained or loss resistance genes. Therefore, I hypothesized that the evolution of IncA/C plasmids has been an ongoing process.

Moreover, certain chromosomal factors may facilitate the common ancestor of node M to acquire the resistant plasmid and to transfer it vertically to offspring strains. The CRISPR/cas system could be the most important part correlated to antibiotic resistance. Palmer et al. demonstrated the inverse correlation between antibiotic resistance and complete CRISPR loci (17). The incomplete CRISPR systems would facilitate the organisms ability to acquire foreign genetic elements and to increase the genome plasticity (16). The diverse accessory gene pools enhance the ability of bacteria to survive in various environments (17). Therefore, it is necessary to perform further whole genome sequencing of node M strains to obtain the complete IncA/C plasmid sequence, to

characterize the antibiotic resistance patterns, to investigate the potential factors facilitating to acquire the plasmids.

References

1. **Thielman NM, Guerrant RL.** 2004. Clinical practice. Acute infectious diarrhea. N Engl J Med **350:**38-47.

2. **Poppe C, Martin L, Muckle A, Archambault M, McEwen S, Weir E.** 2006. Characterization of antimicrobial resistance of *Salmonella* Newport isolated from animals, the environment, and animal food products in Canada. Can J Vet Res **70:**105-114.

3. **Varma JK, Marcus R, Stenzel SA, Hanna SS, Gettner S, Anderson BJ, Hayes T, Shiferaw B, Crume TL, Joyce K, Fullerton KE, Voetsch AC, Angulo FJ.** 2006. Highly resistant *Salmonella* Newport-MDRAmpC transmitted through the domestic US food supply: a FoodNet case-control study of sporadic *Salmonella* Newport infections, 2002-2003. J Infect Dis **194:**222-230.

4. **Gupta A, Fontana J, Crowe C, Bolstorff B, Stout A, Van Duyne S, Hoekstra MP, Whichard JM, Barrett TJ, Angulo FJ.** 2003. Emergence of multidrug-resistant *Salmonella enterica* serotype Newport infections resistant to expanded-spectrum cephalosporins in the United States. J Infect Dis **188:**1707-1716.

5. **Boyd D, Cloeckaert A, Chaslus-Dancla E, Mulvey MR.** 2002. Characterization of variant *Salmonella* genomic island 1 multidrug resistance regions from serovars Typhimurium DT104 and Agona. Antimicrob Agents Chemother **46:**1714-1722.

6. **Welch TJ, Fricke WF, McDermott PF, White DG, Rosso ML, Rasko DA, Mammel MK, Eppinger M, Rosovitz MJ, Wagner D, Rahalison L, Leclerc JE, Hinshaw JM, Lindler LE, Cebula TA, Carniel E, Ravel J.** 2007. Multiple

antimicrobial resistance in plague: an emerging public health risk. PLoS One **2:**e309.

7. **Mulvey MR, Boyd DA, Olson AB, Doublet B, Cloeckaert A.** 2006. The genetics of *Salmonella* genomic island 1. Microbes Infect **8:**1915-1922.

8. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

9. **Vogel J.** 2009. A rough guide to the non-coding RNA world of *Salmonella*. Mol Microbiol **71:**1-11.

10. **Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J.** 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet **4:**e1000163.

11. **Pfeiffer V, Sittka A, Tomer R, Tedin K, Brinkmann V, Vogel J.** 2007. A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. Mol Microbiol **66:**1174-1191.

12. **Padalon-Brauch G, Hershberg R, Elgrably-Weiss M, Baruch K, Rosenshine I, Margalit H, Altuvia S.** 2008. Small RNAs encoded within genetic islands of *Salmonella* typhimurium show host-induced expression and role in virulence. Nucleic Acids Res **36:**1913-1927.

13. **Edwards RA, Schifferli DM, Maloy SR.** 2000. A role for *Salmonella* fimbriae in intraperitoneal infections. Proc Natl Acad Sci U S A **97:**1258-1262.

14. **Porwollik S (ed.).** 2011. *Salmonella*: from Genome to Function. Caister.

15. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167-170.

16. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.

17. **Palmer KL, Gilmore MS.** 2010. Multidrug-resistant *Enterococci* lack CRISPR-cas. MBio **1**.

18. **Darling AE, Mau B, Perna NT.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147.

19. **Goloboff P, Farris S, Nixon K.** 2008. TNT, a program for phylogenetic analysis. Cladistics **24:**774-786.

20. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-2739.

21. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res **35:**W52-57.

22. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.

23. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

24. **Mebrhatu MT, Cenens W, Aertsen A.** 2013. An overview of the domestication and impact of the Salmonella mobilome. Crit Rev Microbiol.

25. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010. Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

26. **Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, White DG, Cebula TA, Ravel J.** 2009. Comparative genomics of the IncA/C multidrug resistance plasmid family. J Bacteriol **191:**4750-4757.

TABLE IV-1. General information of *Salmonella* Newport strains.

Strain Name	Tree Label	Antimicrobial Resistance Profile*	WGS Accession Number	Genome Size (Mbp)
S. Newport SL254	Newport SL254	AMC, AMP, CAZ, CHL, GEN, STR, SUL, TET	ABEN01000000	4.83
CVM 22513	cattle_NC_2003	AMC, AMP, FOX, TIO, AXO, CHL, KAN, STR, SUL, TET	AHTU00000000	4.90
CVM 22425	cattle_AZ_2003	AMC, AMP, FOX, TIO, AXO, CHL, STR, SUL, TET	AHTW00000000	4.93
CVM 22462	canine_AZ_2003	AMC, AMP, FOX, TIO, AXO, CHL, STR, SUL, TET	AHTX00000000	5.02
CVM N1543	ground_beef_GA_2004	AMC, AMP, FOX, TIO, AXO, CHL, STR, SUL, TET	AHTZ00000000	4.89
CVM 21538	chicken_GA	CHL, STR, SUL, TET	AHTV00000000	4.93
CVM 21550	swine_TX	AMC, AMP, FOX, TIO, AXO, CHL, STR, SUL, TET	AHTT00000000	4.92
CVM N18486	ground_turkey_NM_2008	TET	AHTY00000000	4.93

*AMC = Amoxicillin/Clavulanic Acid, AMP = Ampicillin, FOX = Cefoxitin, AXO = Ceftriaxone, CHL = Chloramphenicol, GEN = Gentamicin, KAN = Kanamycin, STR = Streptomycin, SUL = Sulfamethoxazole or Sulfisoxazole , TET = Tetracycline, TIO = Ceftiofur.

TABLE IV-2. Pairwise distance matrix of selected genomes.

	ground_turkey	swine	canine	ground_beef	SL254	chicken	cattle_NC
swine	4194 (20)						
canine	4910 (10)	858 (21)					
ground_beef	4233 (19)	162 (10)	823 (22)				
SL254	4186 (19)	115 (9)	784 (21)	67 (8)			
chicken	4188 (20)	117 (9)	786 (20)	69 (8)	16 (4)		
cattle_NC	4204 (19)	111 (9)	802 (21)	79 (9)	36 (5)	38 (6)	
cattle_AZ	4196 (20)	105 (8)	786 (20)	77 (9)	28 (5)	30 (6)	26 (5)

TABLE IV-3. Identified possible plasmids and resistance genes in selected genomes.

Strain	Identified Complete and Partial Plasmids	Resistance Genes
S. Newport SL254	pSN254, pSN254_3	*floR, tetAR, strAB, sul, sugE, ampC, aadA, aacC, qacE, merABDEPRT*
cattle_NC_2003	pSBardoKan	
cattle_AZ_2003	pSN254_3	
canine_AZ_2003	pSN254, pSN254_3	*floR, tetAR, strAB, merABDEPRT*
ground_beef_GA_2004	pSN254, pSL476_3	*floR, tetAR, strAB, merABDEPRT*
chicken_GA		
swine_TX	pSN254_3	
ground_turkey_NM_2008		
swine_IL_2001		

Table IV-4. Distribution of 106 ncRNA in compared genomes

	SL254	cattle_NC	cattle_AZ	canine	beef	chicken	swine	turkey
isrA	+	+	+	+	+	+	+	+
isrB1	++	++	++	+	+	++	++	+
isrB2	++	++	++	+	+	++	++	+
isrC	+	+	+	+	+	+	+	+
isrD	+	+	+	+	+	+	+	+
isrE	+	+	+	+	+	+	+	+
isrF	+	+	+	+	+	+	+	+
isrG	+	+	+	+	+	+	+	+
isrH1	+	+	+	+	+	+	+	+
isrH2	+	+	+	+	+	+	+	+
isrI	+	+	+	+	+	+	+	+
isrJ	+	+	+	+	+	+	+	+
isrK	+	+	+	+	+	+	+	+
isrL	+	+	+	+	+	+	+	+
isrM	+	+	+	+	+	+	+	+
isrN	+	+	+	+	+	+	+	+
isrO	+	+	+	+	+	+	+	+
isrP	+	+	+	+	+	+	+	+
isrQ	+	+	+	+	+	+	+	+
csrB	+	+	+	+	+	+	+	+

csrC	+	+	+	+	+	+	+	+
cyaR	+	+	+	+	+	+	+	+
dsrA	+	+	+	+	+	+	+	+
gcvB	+	+	+	+	+	+	+	+
glmY	+	+	+	+	+	+	+	+
glmZ	+	+	+	+	+	+	+	+
istR1	+	+	+	+	+	+	+	+
micA	+	+	+	+	+	+	+	+
micC	+	+	+	+	+	+	+	+
micF	+	+	+	+	+	+	+	+
omrA	+	+	+	+	+	+	+	+
omrB	+	+	+	+	+	+	+	+
oxyS	+	+	+	+	+	+	+	+
rprA	+	+	+	+	+	+	+	+
rseX	+	+	+	+	+	+	+	+
rybB	+	+	+	+	+	+	+	+
rydB	+	+	+	+	+	+	+	+
rydC	+	+	+	+	+	+	+	+
ryeB	+	+	+	+	+	+	+	+
ryeC	+	+	+	+	+	+	+	+
ryfA	+	+	+	+	+	+	+	+
rygC	+	+	+	+	+	+	+	+

rygD	+	+	+	+	+	+	+	+
ryhB1	+	+	+	+	+	+	+	+
ryhB2	+	+	+	+	+	+	+	+
sgrS	+	+	+	+	+	+	+	+
spf	+	+	+	+	+	+	+	+
sraF	+	+	+	+	+	+	+	+
sraH	+	+	+	+	+	+	+	+
sraL	+	+	+	+	+	+	+	+
sroB	+	+	+	+	+	+	+	+
sroC	+	+	+	+	+	+	+	+
ssrS	+	+	+	+	+	+	+	+
STnc10	+	+	+	+	+	+	+	+
STnc20	+	+	+	+	+	+	+	+
STnc30	+	+	+	+	+	+	+	+
STnc40	+	+	+	+	+	+	+	+
STnc50	+	+	+	+	+	+	+	+
STnc60	+	+	+	-	+	+	+	+
STnc70	+	+	+	+	+	+	+	+
STnc80	+	+	+	+	+	+	+	+
STnc90	+	+	+	+	+	+	+	+
STnc100	+	+	+	+	+	+	+	+
STnc110	+	+	+	+	+	+	+	+

STnc120	+	+	+	+	+	+	+	+
STnc130	+	+	+	+	+	+	+	+
STnc140	+	+	+	+	+	+	+	+
STnc150	+	+	+	+	+	+	+	+
STnc160	+	+	+	+	+	+	+	+
STnc170	+	+	+	+	+	+	+	+
STnc180	+	+	+	+	+	+	+	+
STnc190	+	+	+	-	+	+	+	+
STnc200	+	+	+	+	+	+	+	+
STnc210	+	+	+	+	+	+	+	+
STnc220	+	+	+	+	+	+	+	+
STnc230	+	+	+	+	+	+	+	+
STnc240	+	+	+	+	+	+	+	+
STnc250	+	+	+	+	+	+	+	+
STnc260	+	+	+	+	+	+	+	+
STnc270	+	+	+	+	+	+	+	+
STnc280	+	+	+	+	+	+	+	+
STnc290	+	+	+	-	+	+	+	+
STnc300	+	+	+	+	+	+	+	+
STnc310	+	+	+	+	+	+	+	+
STnc320	+	+	+	+	+	+	+	+
STnc330	+	+	+	+	+	+	+	+

STnc340	+	+	+	+	+	+	+	+
STnc350	+	+	+	+	+	+	+	+
STnc360	+	+	+	+	+	+	+	+
STnc370	+	+	+	+	+	+	+	+
STnc380	+	+	+	+	+	+	+	+
STnc390	+	+	+	+	+	+	+	+
STnc400	+	+	+	+	+	+	+	+
STnc410	+	+	+	+	+	+	+	+
STnc420	+	+	+	+	+	+	+	+
STnc430	+	+	+	+	+	+	+	+
STnc440	+	+	+	+	+	+	+	+
STnc450	+	+	+	+	+	+	+	+
STnc460	+	+	+	+	+	+	+	+
STnc490	+	+	+	+	+	+	+	+
STnc500	+	+	+	+	+	+	+	+
STnc520	+	+	+	+	+	+	+	+
STnc540	+	+	+	+	+	+	+	+
STnc560	+	+	+	+	+	+	+	+
STnc570	+	+	+	+	+	+	+	+
STnc580	+	+	+	+	+	+	+	+

+: presence; ++: double copies; -: absence

Table IV-5. Point mutations in ncRNAs*

	STnc60	*STnc190*
ground_turkey_NM_2008	A at position 80, 164, 175	T at position 132
ground_beef_GA_2004	T at position 199 and 202	A at position 101
		T at position 110
S. Newport SL254	T at position 199	
cattle_NC_2003	T at position 200	
chicken_GA	A at position 164	
	T at position 200	

*All mutations in Table IV-5 are insertion

FIG IV-1. Phylogenetic tree of compared genomes.

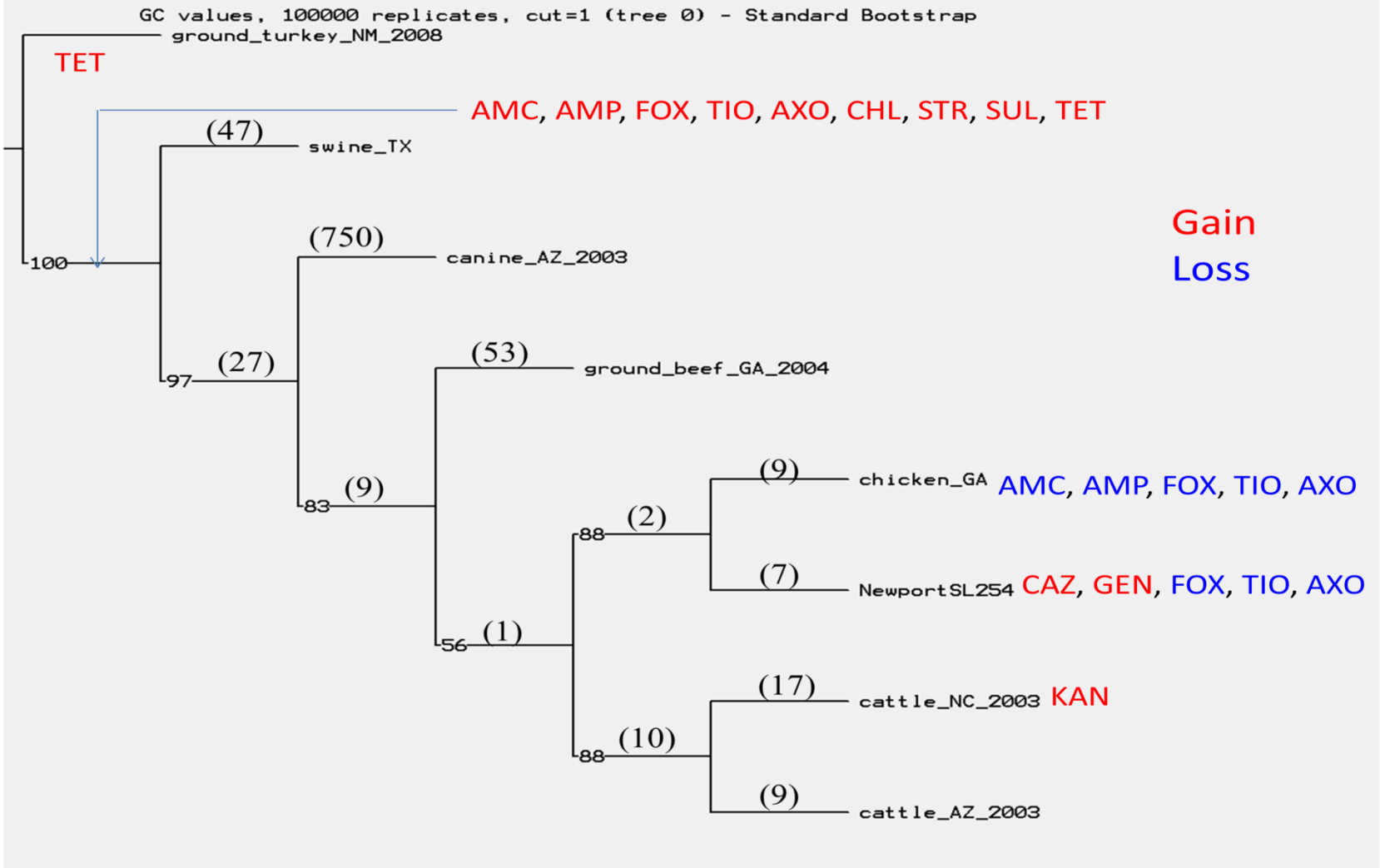
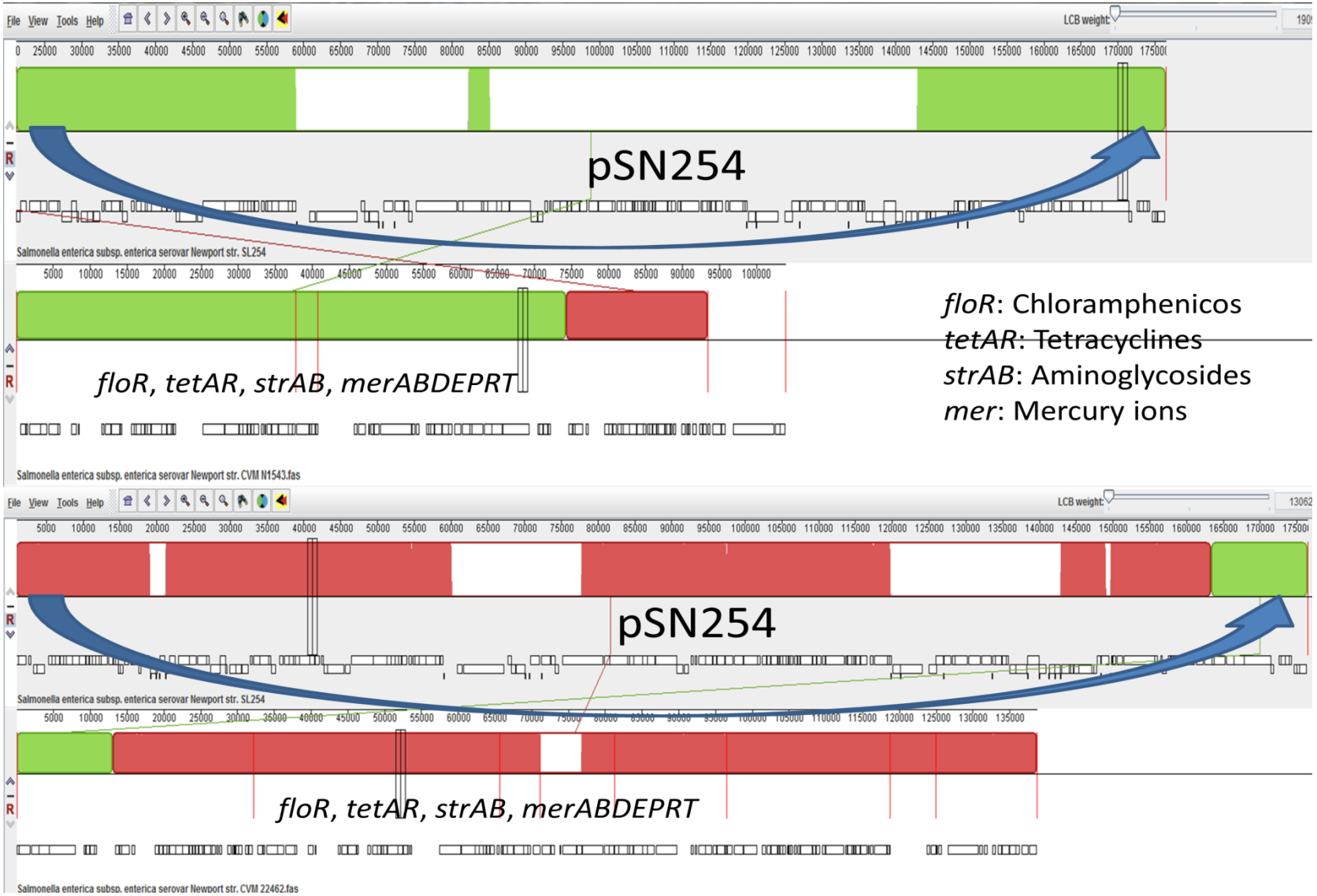


FIG IV-2. Genes in plasmid pSN254 identified in strains canine_AZ_2003 and ground_beef_GA_2004



CHPTER V: SUMMARY AND FUTURE STUDY

Salmonella enterica subspecies *enterica* serotype Newport (*S.* Newport) has been a significant causative agent responsible for foodborne outbreaks in the United States. *S.* Newport has become prevalence since 1996 causing multistate foodborne outbreaks and been isolated from diverse sources and locations. *S.* Newport causes approximate 100,000 illness cases annually in the United States, ranking third most prevalent in more than 1,500 serotypes of *Salmonella enterica* subspecies *enterica*. Moreover, multidrug resistance of *S.* Newport has become one emerging public health burden. Therefore, it is necessary to investigate the evolutionary history, genetic diversity, and multidrug resistance of this clinically important pathogen.

Since the first two complete genomes of pathogenic *Salmonella* were released in 2001 and the fast development of commercial whole genome sequencing (WGS) platforms, WGS has gradually become one essential and important research tool to get insights into the pathogenicity and evolution of *Salmonella*. The availability of WGS data not only provides more comprehensive picture of genetic information of *Salmonella* for research on virulence and host adaptations, but also enables the investigators to locate the potential contamination sources during trace-back epidemiology investigations. WGS has the higher sensitivity and discriminatory power to differentiate close related strains than traditional molecular methods. Moreover, WGS data could be applied in different research fields of food safety, including detection and typing of pathogens, optimize growth and survival strategies.

Chapter II focused on phylogenetic relationship between *S.* Newport lineages II and III. Moreover, several regions in the chromosomes could differentiate the two lineages, such

as regions around *mutS* and *cas* gene clusters in CRISPR system. The findings in chapter II examined the evolutionary history of *S.* Newport from diverse sources and locations from genome-wide view, getting insights into the genetic diversity of different lineages in the same serotype.

A total of 26 *S.* Newport strains isolated from diverse sources and geographic locations were performed pyrosequencing to obtain the draft genomes. Two publicly available *S.* Newport genomes were selected as reference genomes of lineages II and III, respectively. Another 15 genomes belonging to different serotypes were selected as outgroup genomes to determine the phylogenetic relationship between compared genomes. Lineages II and III displayed diverse relationship with each other, proving that *S.* Newport was polyphyletic with extensive diversity. Lineages II and III were separated by other serotypes, indicating that both two lineages evolved largely independently after they diverged at the early stage of *S.* Newport evolution. Moreover, *S.* Newport harbored a clear geographic structure, meaning that the strains from Asia were decoupled from those from the Americas. Lineage II showed more diversity structure than lineage III and consisted of three subgroups.

The region around *mutS* was examined because recombination events happened frequently at this location, which could be used as biomarkers to differentiate lineages II and III. This region in lineages II or III shared common contents with different outgroup genomes. For example, all 15 outgroup genomes shared gene cluster 1 between *invH* and *mutS* with lineage II while lineage III possessed gene cluster 2 at the same location. Similarly, *cas* gene cluster also differentiate lineages II and III.

Therefore, *S.* Newport had a clear geographic structure and lineages II and III displayed extensive genetic diversity compared to each other. The polyphyletic structure of *S.* Newport suggested that a large gene pool of this serotype may exist. It is necessary to collect more *S.* Newport strains to identify novel subgroups of both lineages, especially those isolated from different geographic locations. Moreover, as more WGS data became available, new biomarkers could be determined to distinguish various lineages or subgroups of *S.* Newport.

Chapter III focused on diversity of important *Salmonella* pathogenicity islands (SPIs). Because of the important roles of these gene clusters in invasion activity, interactions between hosts and cells, and secretion of virulent factors, the diversity of SPIs may cause differences in pathogenicity of different genomes. Here, SPI-5 and SPI-6 were chosen to investigate their genetic diversity.

A total of 28 *S.* Newport from diverse sources and locations were selected. Another 11 genomes of other serotypes were selected as outgroup genomes. A total of 146 single nucleotide polymorphisms of SPI-5 were identified. The phylogenetic tree of genes in SPI-5 indicated that lineages II and III showed divergent relationship and were separated by other outgroup genomes. Moreover, two different large insertions with over 40 kb were determined in SPI-5 in certain genomes, indicating that the evolution of SPI-5 was an ongoing process and horizontal gene transfer or recombination events played important roles. Furthermore, because insertion 1 was disseminated in different genomes shared common ancestor, it may be compatible to the overall chromosome and play important roles in virulence activities.

SPI-6, containing type VI secretion systems (T6SS), was identified in all genomes except the Asian strains in subgroup IIA. The absence of SPI-6 or T6SS in these strains indicated that they may have different pathogenicity capacity or host adaptations. Of those carrying SPI-6, the phylogenetic analysis showed that lineages II and III were separated by outgroup genomes. Moreover, the presence of *tcf* fimbrial in certain genomes suggested that this region could be hot spot for horizontal gene transfer and recombination events. Although the genetic diversity were identified in both SPI-5 and SPI-6, experiment work is required to prove the hypothesis that the mutations including indels and SNPs would result in the differences in pathogenicity of these genomes.

In chapter IV, seven MDR *S.* Newport in subgroup IIC were investigated to identify their common and specific genetic information. The phylogenetic analysis suggested that these seven genomes had close relationship with each other and originated from common ancestor. Therefore, it is necessary to determine the possible common genomic background information among these genomes.

Complete and remnant plasmid sequences were determined in the draft genomes data, including *floR* (chloramphenicols), *tetAR* (tetracyclines), *strAB* (aminoglycosides), and *merABDEPRT* (mercury ions). For example, plasmid pSN254, which was responsible for ACSSuT resistance type, existed in two genomes. Moreover, all genomes in this group contained pseudogene *cas3* in CRISPR/cas system. Thus, the incomplete CRISPR/cas system may facilitate these strains to acquire foreign elements and to facilitate the survival of these strains in different environments. Further chromosomes and plasmids sequencing are necessary to classify resistance plasmids in these genomes and to investigate the potential chromosomal factors to facilitate their acquisition of foreign

elements. In addition, it is necessary to investigate more MDR *S.* Newport strains to check their phylogenies and to identify the elements responsible to antimicrobial resistance.

MASTER REFERNECES

References in Chapter I

1. **Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, Petty NK, Kingsley RA, Baumler AJ, Nuccio SP, Contreras I, Santiviago CA, Maskell D, Barrow P, Humphrey T, Nastasi A, Roberts M, Frankel G, Parkhill J, Dougan G, Thomson NR.** 2011. *Salmonella bongori* provides insights into the evolution of the Salmonellae. PLoS Pathog **7:**e1002191.
2. **Doolittle RF, Feng DF, Tsang S, Cho G, Little E.** 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271:**470-477.
3. **Reeves MW, Evins GM, Heiba AA, Plikaytis BD, Farmer JJ, 3rd.** 1989. Clonal nature of *Salmonella* typhi and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. J Clin Microbiol **27:**313-320.
4. **Humphrey T.** 2004. *Salmonella*, stress responses and food safety. Nat Rev Microbiol **2:**504-509.
5. **Hickman-Brenner FW, Stubbs AD, Farmer JJ, 3rd.** 1991. Phage typing of *Salmonella* Enteritidis in the United States. J Clin Microbiol **29:**2817-2823.
6. **De Lappe N, Doran G, O'Connor J, O'Hare C, Cormican M.** 2009. Characterization of bacteriophages used in the *Salmonella enterica* serovar Enteritidis phage-typing scheme. J Med Microbiol **58:**86-93.

7. **Harbottle H, White DG, McDermott PF, Walker RD, Zhao S.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J Clin Microbiol **44:**2449-2457.

8. **Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW.** 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One **8:**e55254.

9. **Beltran P, Musser JM, Helmuth R, Farmer JJ, 3rd, Frerichs WM, Wachsmuth IK, Ferris K, McWhorter AC, Wells JG, Cravioto A, et al.** 1988. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S.* choleraesuis, *S.* derby, *S.* dublin, *S.* enteritidis, *S.* heidelberg, *S.* infantis, *S.* newport, and *S.* typhimurium. Proc Natl Acad Sci U S A **85:**7753-7757.

10. **Kotetishvili M, Stine OC, Kreger A, Morris JG, Jr., Sulakvelidze A.** 2002. Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. J Clin Microbiol **40:**1626-1635.

11. **Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG.** 2011. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. Appl Environ Microbiol **77:**1946-1956.

12. **Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C, Weill FX.** 2012. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One **7:**e36995.

13. **Bell RL, Cao G, Meng J, Allard MW, Keys CE, Hill T, Ottesen AR, Brown EW.** 2012. *Salmonella* Newport Contamination of Produce: Ecological, Genetic, and Epidemiological Aspects, p. In press, *Salmonella*: Classification, Genetics and Disease. Nova Publishers, Hauppauge NY.

14. **Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao G, Meng J, Stones R.** 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med **364:**981-982.

15. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-625.

16. **Voetsch AC, Van Gilder TJ, Angulo FJ, Farley MM, Shallow S, Marcus R, Cieslak PR, Deneen VC, Tauxe RV.** 2004. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin Infect Dis **38 Suppl 3:**S127-134.

17. **Haraga A, Ohlson MB, Miller SI.** 2008. Salmonellae interplay with host cells. Nat Rev Microbiol **6:**53-66.

18. **Woodward DL, Khakhria R, Johnson WM.** 1997. Human salmonellosis associated with exotic pets. J Clin Microbiol **35:**2786-2790.

19. **Sabbagh SC, Forest CG, Lepage C, Leclerc JM, Daigle F.** 2010. So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. FEMS Microbiol Lett **305:**1-13.

20. **CDC.** 2009. *Salmonella* Surveillance: Annual Summary. Atlanta, Georgia: US Department of Health and Human Services, CDC.

21. **Garcia-del Portillo F, Foster JW, Finlay BB.** 1993. Role of acid tolerance response genes in *Salmonella* Typhimurium virulence. Infect Immun **61:**4489-4492.

22. **Michetti P, Mahan MJ, Slauch JM, Mekalanos JJ, Neutra MR.** 1992. Monoclonal secretory immunoglobulin A protects mice against oral challenge with the invasive pathogen *Salmonella* Typhimurium. Infect Immun **60:**1786-1792.

23. **Francis CL, Starnbach MN, Falkow S.** 1992. Morphological and cytoskeletal changes in epithelial cells occur immediately upon interaction with *Salmonella* Typhimurium grown under low-oxygen conditions. Mol Microbiol **6:**3077-3087.

24. **Greene SK, Daly ER, Talbot EA, Demma LJ, Holzbauer S, Patel NJ, Hill TA, Walderhaug MO, Hoekstra RM, Lynch MF, Painter JA.** 2008. Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. Epidemiol Infect **136:**157-165.

25. 2011. Vital signs: incidence and trends of infection with pathogens transmitted commonly through food--foodborne diseases active surveillance network, 10 U.S. sites, 1996-2010. MMWR Morb Mortal Wkly Rep **60:**749-755.

26. **Van Beneden CA, Keene WE, Strang RA, Werker DH, King AS, Mahon B, Hedberg K, Bell A, Kelly MT, Balan VK, Mac Kenzie WR, Fleming D.** 1999. Multinational outbreak of *Salmonella enterica* serotype Newport infections due to contaminated alfalfa sprouts. JAMA **281:**158-162.

27. **Sivapalasingam S, Barrett E, Kimura A, Van Duyne S, De Witt W, Ying M, Frisch A, Phan Q, Gould E, Shillam P, Reddy V, Cooper T, Hoekstra M, Higgins C, Sanders JP, Tauxe RV, Slutsker L.** 2003. A multistate outbreak of *Salmonella enterica* Serotype Newport infection linked to mango consumption: impact of water-dip disinfestation technology. Clin Infect Dis **37:**1585-1590.

28. **Kirk MD, Little CL, Lem M, Fyfe M, Genobile D, Tan A, Threlfall J, Paccagnella A, Lightfoot D, Lyi H, McIntyre L, Ward L, Brown DJ, Surnam S, Fisher IS.** 2004. An outbreak due to peanuts in their shell caused by *Salmonella enterica* serotypes Stanley and Newport--sharing molecular information to solve international outbreaks. Epidemiol Infect **132:**571-577.

29. **Schneider JL, White PL, Weiss J, Norton D, Lidgard J, Gould LH, Yee B, Vugia DJ, Mohle-Boetani J.** 2011. Multistate outbreak of multidrug-resistant *Salmonella* newport infections associated with ground beef, October to December 2007. J Food Prot **74:**1315-1319.

30. **Jackson BR, Griffin PM, Cole D, Walsh KA, Chai SJ.** 2013. Outbreak-associated *Salmonella enterica* Serotypes and Food Commodities, United States, 1998-2008. Emerg Infect Dis **19:**1239-1244.

31. **Zansky S, Wallace B, Schoonmaker-Bopp D, Smith P, Ramsey F, Painter J, Gupta A, Kalluri P, Noviello S.** 2002. From the Centers for Disease Control and Prevention. Outbreak of multi-drug resistant *Salmonella* Newport--United States, January-April 2002. JAMA **288:**951-953.

32. **Varma JK, Marcus R, Stenzel SA, Hanna SS, Gettner S, Anderson BJ, Hayes T, Shiferaw B, Crume TL, Joyce K, Fullerton KE, Voetsch AC, Angulo FJ.** 2006. Highly resistant *Salmonella* Newport-MDRAmpC transmitted through the domestic US food supply: a FoodNet case-control study of sporadic *Salmonella* Newport infections, 2002-2003. J Infect Dis **194:**222-230.

33. **Gupta A, Fontana J, Crowe C, Bolstorff B, Stout A, Van Duyne S, Hoekstra MP, Whichard JM, Barrett TJ, Angulo FJ.** 2003. Emergence of multidrug-resistant *Salmonella enterica* serotype Newport infections resistant to expanded-spectrum cephalosporins in the United States. J Infect Dis **188:**1707-1716.

34. **Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG.** 2001. Complete genome sequence of a

multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848-852.

35. **McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK.** 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature **413:**852-856.

36. **Mills DM, Bajaj V, Lee CA.** 1995. A 40 kb chromosomal fragment encoding *Salmonella* typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome. Mol Microbiol **15:**749-759.

37. **Galan JE.** 1999. Interaction of *Salmonella* with host cells through the centisome 63 type III secretion system. Curr Opin Microbiol **2:**46-50.

38. **Hu Q, Coburn B, Deng W, Li Y, Shi X, Lan Q, Wang B, Coombes BK, Finlay BB.** 2008. *Salmonella enterica* serovar Senftenberg human clinical isolates lacking SPI-1. J Clin Microbiol **46:**1330-1336.

39. **Brown EW, Mammel MK, LeClerc JE, Cebula TA.** 2003. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens. Proc Natl Acad Sci U S A **100:**15676-15681.

40. **Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE.** 1998. Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. Mol Microbiol **29:**883-891.

41. **Hensel M.** 2004. Evolution of pathogenicity islands of *Salmonella enterica*. Int J Med Microbiol **294:**95-102.

42. **Knodler LA, Celli J, Hardt WD, Vallance BA, Yip C, Finlay BB.** 2002. *Salmonella* effectors within a single pathogenicity island are differentially expressed and translocated by separate type III secretion systems. Mol Microbiol **43:**1089-1103.

43. **Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Lofdahl S.** 1999. Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. Mol Microbiol **33:**612-622.

44. **Blondel CJ, Jimenez JC, Contreras I, Santiviago CA.** 2009. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. BMC Genomics **10:**354.

45. **Jani AJ, Cotter PA.** 2010. Type VI secretion: not just for pathogenesis anymore. Cell Host Microbe **8:**2-6.

46. **Sukhnanand S, Alcaine S, Warnick LD, Su WL, Hof J, Craver MP, McDonough P, Boor KJ, Wiedmann M.** 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella* enterica serotypes. J Clin Microbiol **43:**3688-3698.

47. **Alcaine SD, Soyer Y, Warnick LD, Su WL, Sukhnanand S, Richards J, Fortes ED, McDonough P, Root TP, Dumas NB, Grohn Y, Wiedmann M.** 2006. Multilocus sequence typing supports the hypothesis that cow- and human-associated *Salmonella* isolates represent distinct and overlapping populations. Appl Environ Microbiol **72:**7575-7585.

48. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010.

Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

49. **Begley M, Hill C.** 2010. Food safety: What can we learn from genomics? Annu Rev Food Sci Technol **1:**341-361.

50. **Mardis ER.** 2013. Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto Calif) **6:**287-303.

References in Chapter II

1. **Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-625.

2. **Voetsch AC, Van Gilder TJ, Angulo FJ, Farley MM, Shallow S, Marcus R, Cieslak PR, Deneen VC, Tauxe RV.** 2004. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin Infect Dis **38 Suppl 3:**S127-134.

3. **CDC.** 2006. *Salmonella* Surveillance: Annual Summary. Atlanta, Georgia: US Department of Health and Human Services, CDC.

4. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.

5. **Patrick AD GF-X.** 2007. Antigenic formulae of the Salmonella serovars, 9th ed.

6. **Attorney S.** 2009. Multi-state *Salmonella* Newport outbreak leads to huge ground beef recall.

7. **Attorney S.** 2011. Sprout *Salmonella* Newport outbreak sickens six in Oregon and Washington.

8. **CDC.** 2010. Investigation update: multi-state outbreak of human *Salmonella* Newport infections linked to raw alfalfa sprouts. Atlanta, Georgia: US Department of Health and Human Services, CDC.

9. **Bell RL, Cao G, Meng J, Allard MW, Keys CE, Hill T, Ottesen AR, Brown EW.** 2012. *Salmonella* Newport Contamination of Produce: Ecological, Genetic, and Epidemiological Aspects, p. In press, *Salmonella*: Classification, Genetics and Disease. Nova Publishers, Hauppauge NY.

10. **Beltran P, Musser JM, Helmuth R, Farmer JJ, 3rd, Frerichs WM, Wachsmuth IK, Ferris K, McWhorter AC, Wells JG, Cravioto A, et al.** 1988. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S.* choleraesuis, *S.* derby, *S.* dublin, *S.* enteritidis, *S.* heidelberg, *S.* infantis, *S.* newport, and *S.* typhimurium. Proc Natl Acad Sci U S A **85:**7753-7757.

11. **Harbottle H, White DG, McDermott PF, Walker RD, Zhao S.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J Clin Microbiol **44:**2449-2457.

12. **Sukhnanand S, Alcaine S, Warnick LD, Su WL, Hof J, Craver MP, McDonough P, Boor KJ, Wiedmann M.** 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella* enterica serotypes. J Clin Microbiol **43:**3688-3698.

13. **Torpdahl M, Skov MN, Sandvang D, Baggesen DL.** 2005. Genotypic characterization of *Salmonella* by multilocus sequence typing, pulsed-field gel electrophoresis and amplified fragment length polymorphism. J Microbiol Methods **63:**173-184.

14. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010. Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

15. **Didelot X, Achtman M, Parkhill J, Thomson NR, Falush D.** 2007. A bimodal pattern of relatedness between the *Salmonella* Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination? Genome Res **17:**61-68.

16. **Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, Leclerc JE, Ravel J, Cebula TA.** 2011. Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. J Bacteriol **193:**3556-3568.

17. **Brown EW, Mammel MK, LeClerc JE, Cebula TA.** 2003. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens. Proc Natl Acad Sci U S A **100:**15676-15681.

18. **Mills DM, Bajaj V, Lee CA.** 1995. A 40 kb chromosomal fragment encoding *Salmonella* typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome. Mol Microbiol **15:**749-759.

19. **Hu Q, Coburn B, Deng W, Li Y, Shi X, Lan Q, Wang B, Coombes BK, Finlay BB.** 2008. *Salmonella enterica* serovar Senftenberg human clinical isolates lacking SPI-1. J Clin Microbiol **46:**1330-1336.

20. **Sorek R, Kunin V, Hugenholtz P.** 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol **6:**181-186.

21. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A.** 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol **169:**5429-5433.

22. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

23. **Deveau H, Garneau JE, Moineau S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol **64:**475-493.

24. **Vale PF, Little TJ.** 2010. CRISPR-mediated phage resistance and the ghost of coevolution past. Proc Biol Sci **277:**2097-2103.

25. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.

26. **McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller**

W, Stoneking T, Nhan M, Waterston R, Wilson RK. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature **413:**852-856.

27. **Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848-852.

28. **Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK.** 2011. The origin of the Haitian cholera outbreak strain. N Engl J Med **364:**33-42.

29. **Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao G, Meng J, Stones R.** 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med **364:**981-982.

30. **Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD.** 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science **327:**469-474.

31. **Maki DG.** 2009. Coming to grips with foodborne infection--peanut butter, peppers, and nationwide *Salmonella* outbreaks. N Engl J Med **360:**949-953.

32. **Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW.** 2012. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **13:**32.

33. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

34. **den Bakker HC, Moreno Switt AI, Govoni G, Cummings CA, Ranieri ML, Degoricija L, Hoelzer K, Rodriguez-Rivera LD, Brown S, Bolchacova E, Furtado MR, Wiedmann M.** 2011. Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*. BMC Genomics **12:**425.

35. **Jacobsen A, Hendriksen RS, Aaresturp FM, Ussery DW, Friis C.** 2011. The *Salmonella* enterica pan-genome. Microb Ecol **62:**487-504.

36. **Klimke W, Agarwala R, Badretdin A, Chetvernin S, Ciufo S, Fedorov B, Kiryutin B, O'Neill K, Resch W, Resenchuk S, Schafer S, Tolstoy I, Tatusova T.** 2009. The National Center for Biotechnology Information's Protein Clusters Database. Nucleic Acids Res **37:**D216-223.

37. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.

38. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

39. **Goloboff P, Farris S, Nixon K.** 2008. TNT, a program for phylogenetic analysis. Cladistics **24:**774-786.

40. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221-224.

41. **Didelot X, Falush D.** 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics **175:**1251-1266.

42. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-2739.

43. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460-2461.

44. **Dunne EF, Fey PD, Kludt P, Reporter R, Mostashari F, Shillam P, Wicklund J, Miller C, Holland B, Stamey K, Barrett TJ, Rasheed JK, Tenover FC, Ribot EM, Angulo FJ.** 2000. Emergence of domestically acquired ceftriaxone-resistant *Salmonella* infections associated with AmpC beta-lactamase. JAMA **284:**3151-3156.

45. **Fey PD, Safranek TJ, Rupp ME, Dunne EF, Ribot E, Iwen PC, Bradford PA, Angulo FJ, Hinrichs SH.** 2000. Ceftriaxone-resistant *Salmonella* infection acquired by a child from cattle. N Engl J Med **342:**1242-1249.

46. **Park AK, Kim H, Jin HJ.** 2010. Phylogenetic analysis of rRNA methyltransferases, Erm and KsgA, as related to antibiotic resistance. FEMS Microbiol Lett **309:**151-162.

47. **Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D, Donnelly P.** 2011. Recombination and population structure in *Salmonella enterica*. PLoS Genet **7:**e1002191.

48. **Welch TJ, Fricke WF, McDermott PF, White DG, Rosso ML, Rasko DA, Mammel MK, Eppinger M, Rosovitz MJ, Wagner D, Rahalison L, Leclerc JE, Hinshaw JM, Lindler LE, Cebula TA, Carniel E, Ravel J.** 2007. Multiple antimicrobial resistance in plague: an emerging public health risk. PLoS One **2:**e309.

49. **Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, White DG, Cebula TA, Ravel J.** 2009. Comparative genomics of the IncA/C multidrug resistance plasmid family. J Bacteriol **191:**4750-4757.

50. **Matsumura K, Furukawa S, Ogihara H, Morinaga Y.** 2011. Roles of multidrug efflux pumps on the biofilm formation of *Escherichia coli* K-12. Biocontrol Sci **16:**69-72.

51. **Yamasaki S, Nagasawa S, Hayashi-Nishino M, Yamaguchi A, Nishino K.** 2011. AcrA dependency of the AcrD efflux pump in *Salmonella enterica* serovar Typhimurium. J Antibiot (Tokyo) **64:**433-437.

52. **Lama A, Pane-Farre J, Chon T, Wiersma AM, Sit CS, Vederas JC, Hecker M, Nakano MM.** 2012. Response of Methicillin-Resistant *Staphylococcus aureus* to Amicoumacin A. PLoS One **7:**e34037.

53. **Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, Petty NK, Kingsley RA, Baumler AJ, Nuccio SP, Contreras I, Santiviago CA, Maskell D, Barrow P, Humphrey T, Nastasi A, Roberts M, Frankel G, Parkhill J, Dougan G, Thomson NR.** 2011. *Salmonella bongori* provides insights into the evolution of the Salmonellae. PLoS Pathog **7:**e1002191.

54. **Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol **60:**1136-1151.

55. **Orsi RH, Sun Q, Wiedmann M.** 2008. Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*. BMC Evol Biol **8:**233.

56. **Lefebure T, Stanhope MJ.** 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol **8:**R71.

57. **Osborne SE, Tuinema BR, Mok MC, Lau PS, Bui NK, Tomljenovic-Berube AM, Vollmer W, Zhang K, Junop M, Coombes BK.** 2012. Characterization of DalS, an ATP-binding cassette transporter for D-alanine, and its role in pathogenesis in *Salmonella enterica*. J Biol Chem.

58. **Chang G.** 2003. Multidrug resistance ABC transporters. FEBS Lett **555:**102-105.

59. **Selander RK, Beltran P, Smith NH, Helmuth R, Rubin FA, Kopecko DJ, Ferris K, Tall BD, Cravioto A, Musser JM.** 1990. Evolutionary genetic

relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. Infect Immun **58:**2262-2275.

60. **Didelot X, Maiden MC.** 2010. Impact of recombination on bacterial evolution. Trends Microbiol **18:**315-322.

61. **Porwollik S (ed.).** 2011. *Salmonella*: from Genome to Function. Caister.

62. **Brown EW, Kotewicz ML, Cebula TA.** 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. Mol Phylogenet Evol **24:**102-120.

63. **Groisman EA, Ochman H.** 1997. How Salmonella became a pathogen. Trends Microbiol **5:**343-349.

References in Chapter III

1. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17:**7-15.

2. **CDC.** 2006. *Salmonella* Surveillance: Annual Summary.

3. **Greene SK, Daly ER, Talbot EA, Demma LJ, Holzbauer S, Patel NJ, Hill TA, Walderhaug MO, Hoekstra RM, Lynch MF, Painter JA.** 2008. Recurrent multistate outbreak of *Salmonella* Newport associated with tomatoes from contaminated fields, 2005. Epidemiol Infect **136:**157-165.

4. **Schneider JL, White PL, Weiss J, Norton D, Lidgard J, Gould LH, Yee B, Vugia DJ, Mohle-Boetani J.** 2011. Multistate outbreak of multidrug-resistant *Salmonella* newport infections associated with ground beef, October to December 2007. J Food Prot **74:**1315-1319.

5. **Bell RL, Cao G, Meng J, Allard MW, Keys CE, Hill T, Ottesen AR, Brown EW.** 2012. *Salmonella* Newport Contamination of Produce: Ecological, Genetic, and Epidemiological Aspects, p. In press, *Salmonella*: Classification, Genetics and Disease. Nova Publishers, Hauppauge NY.

6. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010. Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

7. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

8. **Sabbagh SC, Forest CG, Lepage C, Leclerc JM, Daigle F.** 2010. So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. FEMS Microbiol Lett **305:**1-13.

9. **Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE.** 1998. Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. Mol Microbiol **29:**883-891.

10. **Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L.** 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. Mol Microbiol **39:**260-271.

11. **Knodler LA, Celli J, Hardt WD, Vallance BA, Yip C, Finlay BB.** 2002. *Salmonella* effectors within a single pathogenicity island are differentially

expressed and translocated by separate type III secretion systems. Mol Microbiol **43:**1089-1103.

12. **Hensel M.** 2004. Evolution of pathogenicity islands of *Salmonella enterica*. Int J Med Microbiol **294:**95-102.

13. **Rychlik I, Karasova D, Sebkova A, Volf J, Sisak F, Havlickova H, Kummer V, Imre A, Szmolka A, Nagy B.** 2009. Virulence potential of five major pathogenicity islands (SPI-1 to SPI-5) of *Salmonella enterica* serovar Enteritidis for chickens. BMC Microbiol **9:**268.

14. **Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Lofdahl S.** 1999. Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. Mol Microbiol **33:**612-622.

15. **Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella* enterica serovar Typhi CT18. Nature **413:**848-852.

16. **Schwarz S, Hood RD, Mougous JD.** 2010. What is type VI secretion doing in all those bugs? Trends Microbiol **18:**531-537.

17. **Blondel CJ, Jimenez JC, Contreras I, Santiviago CA.** 2009. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. BMC Genomics **10:**354.

18. **Jani AJ, Cotter PA.** 2010. Type VI secretion: not just for pathogenesis anymore. Cell Host Microbe **8:**2-6.

19. **Pukatzki S, McAuley SB, Miyata ST.** 2009. The type VI secretion system: translocation of effectors and effector-domains. Curr Opin Microbiol **12:**11-17.

20. **Hood RD, Singh P, Hsu F, Guvener T, Carl MA, Trinidad RR, Silverman JM, Ohlson BB, Hicks KG, Plemel RL, Li M, Schwarz S, Wang WY, Merz AJ, Goodlett DR, Mougous JD.** 2010. A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. Cell Host Microbe **7:**25-37.

21. **Folkesson A, Lofdahl S, Normark S.** 2002. The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. Res Microbiol **153:**537-545.

22. **Humphries AD, Raffatellu M, Winter S, Weening EH, Kingsley RA, Droleskey R, Zhang S, Figueiredo J, Khare S, Nunes J, Adams LG, Tsolis RM, Baumler AJ.** 2003. The use of flow cytometry to detect expression of subunits encoded by 11 Salmonella enterica serotype Typhimurium fimbrial operons. Mol Microbiol **48:**1357-1376.

23. **Cao G, Zhao S, Strain E, Luo Y, Timme R, Wang C, Brown E, Meng J, Allard M.** 2012. Draft genome sequences of eight *Salmonella enterica* serotype Newport strains from diverse hosts and locations. J Bacteriol **194:**5146.

24. **Lienau EK, Blazar JM, Wang C, Brown EW, Stones R, Musser S, Allard MW.** 2013. Phylogenomic Analysis Identifies Gene Gains That Define *Salmonella enterica* Subspecies I. PLoS One **8:**e76821.

25. **Timme RE, Pettengill JB, Allard MW, Strain E, Barrangou R, Wehnes C, Van Kessel JS, Karns JS, Musser SM, Brown EW.** 2013. Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. Genome Biol Evol **5:**2109-2123.

26. **Goloboff P, Farris S, Nixon K.** 2008. TNT, a program for phylogenetic analysis. Cladistics **24:**774-786.

27. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

28. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.

29. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

30. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-2739.

31. **Scheurwater E, Reid CW, Clarke AJ.** 2008. Lytic transglycosylases: bacterial space-making autolysins. Int J Biochem Cell Biol **40:**586-591.

32. **Boyd EF, Davis BM, Hochhut B.** 2001. Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. Trends Microbiol **9:**137-144.

33. **Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW.** 2012. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **13:**32.

34. **Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW.** 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One **8:**e55254.

35. **Soyer Y, Orsi RH, Rodriguez-Rivera LD, Sun Q, Wiedmann M.** 2009. Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes. BMC Evol Biol **9:**264.

36. **Zhulin IB, Nikolskaya AN, Galperin MY.** 2003. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. J Bacteriol **185:**285-294.

37. **Luo Y, Bertero MG, Frey EA, Pfuetzner RA, Wenk MR, Creagh L, Marcus SL, Lim D, Sicheri F, Kay C, Haynes C, Finlay BB, Strynadka NC.** 2001. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. Nat Struct Biol **8:**1031-1036.

38. **Hill CW.** 1999. Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. Res Microbiol **150:**665-674.

39. **Porwollik S (ed.).** 2011. *Salmonella*: from Genome to Function. Caister.

40. **Yue M, Rankin SC, Blanchet RT, Nulton JD, Edwards RA, Schifferli DM.** 2012. Diversification of the *Salmonella* fimbriae: a model of macro- and microevolution. PLoS One **7:**e38596.

References in Chapter IV

1. **Thielman NM, Guerrant RL.** 2004. Clinical practice. Acute infectious diarrhea. N Engl J Med **350:**38-47.

2. **Poppe C, Martin L, Muckle A, Archambault M, McEwen S, Weir E.** 2006. Characterization of antimicrobial resistance of *Salmonella* Newport isolated from animals, the environment, and animal food products in Canada. Can J Vet Res **70:**105-114.

3. **Varma JK, Marcus R, Stenzel SA, Hanna SS, Gettner S, Anderson BJ, Hayes T, Shiferaw B, Crume TL, Joyce K, Fullerton KE, Voetsch AC, Angulo FJ.** 2006. Highly resistant *Salmonella* Newport-MDRAmpC transmitted through the domestic US food supply: a FoodNet case-control study of sporadic *Salmonella* Newport infections, 2002-2003. J Infect Dis **194:**222-230.

4. **Gupta A, Fontana J, Crowe C, Bolstorff B, Stout A, Van Duyne S, Hoekstra MP, Whichard JM, Barrett TJ, Angulo FJ.** 2003. Emergence of multidrug-resistant *Salmonella enterica* serotype Newport infections resistant to expanded-spectrum cephalosporins in the United States. J Infect Dis **188:**1707-1716.

5. **Boyd D, Cloeckaert A, Chaslus-Dancla E, Mulvey MR.** 2002. Characterization of variant *Salmonella* genomic island 1 multidrug resistance regions from serovars Typhimurium DT104 and Agona. Antimicrob Agents Chemother **46:**1714-1722.

6. **Welch TJ, Fricke WF, McDermott PF, White DG, Rosso ML, Rasko DA, Mammel MK, Eppinger M, Rosovitz MJ, Wagner D, Rahalison L, Leclerc JE, Hinshaw JM, Lindler LE, Cebula TA, Carniel E, Ravel J.** 2007. Multiple antimicrobial resistance in plague: an emerging public health risk. PLoS One **2:**e309.

7. **Mulvey MR, Boyd DA, Olson AB, Doublet B, Cloeckaert A.** 2006. The genetics of *Salmonella* genomic island 1. Microbes Infect **8:**1915-1922.

8. **Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E, Allard M.** 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One **8:**e55687.

9. **Vogel J.** 2009. A rough guide to the non-coding RNA world of *Salmonella*. Mol Microbiol **71:**1-11.

10. **Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J.** 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet **4:**e1000163.

11. **Pfeiffer V, Sittka A, Tomer R, Tedin K, Brinkmann V, Vogel J.** 2007. A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. Mol Microbiol **66:**1174-1191.

12. **Padalon-Brauch G, Hershberg R, Elgrably-Weiss M, Baruch K, Rosenshine I, Margalit H, Altuvia S.** 2008. Small RNAs encoded within genetic islands of

Salmonella typhimurium show host-induced expression and role in virulence. Nucleic Acids Res **36:**1913-1927.

13. **Edwards RA, Schifferli DM, Maloy SR.** 2000. A role for *Salmonella* fimbriae in intraperitoneal infections. Proc Natl Acad Sci U S A **97:**1258-1262.
14. **Porwollik S (ed.).** 2011. *Salmonella*: from Genome to Function. Caister.
15. **Horvath P, Barrangou R.** 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science **327:**167-170.
16. **Touchon M, Rocha EP.** 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One **5:**e11126.
17. **Palmer KL, Gilmore MS.** 2010. Multidrug-resistant *Enterococci* lack CRISPR-cas. MBio **1**.
18. **Darling AE, Mau B, Perna NT.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147.
19. **Goloboff P, Farris S, Nixon K.** 2008. TNT, a program for phylogenetic analysis. Cladistics **24:**774-786.
20. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-2739.
21. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res **35:**W52-57.

22. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.

23. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

24. **Mebrhatu MT, Cenens W, Aertsen A.** 2013. An overview of the domestication and impact of the Salmonella mobilome. Crit Rev Microbiol.

25. **Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P, Achtman M.** 2010. Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol **192:**6465-6476.

26. **Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, White DG, Cebula TA, Ravel J.** 2009. Comparative genomics of the IncA/C multidrug resistance plasmid family. J Bacteriol **191:**4750-4757.